

(TM)

UNIVERSITY OF EDINBURGH, U.K.  
Distribution rights by Oxford Molecular Ltd

Sequence: 1 MPAPRCRAVRSLLRSYRE.....TALEAAANPALPSDFKTILD 1132

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	134	1-6	509	1	P31060	3.44e-01
2	125	1-5	402	26	P36855	1.45e+00
3	125	1-5	402	10	R50238	1.45e+00
4	125	1-5	402	10	R50238	1.45e+00
5	125	1-5	402	11	R50377	1.45e+00
6	125	1-5	402	9	R57973	1.45e+00
7	125	1-5	402	5	R27291	1.45e+00
8	125	1-5	402	9	R46732	1.45e+00
9	125	1-5	402	22	W16365	1.45e+00
10	125	1-5	402	8	R44759	1.45e+00
11	125	1-5	402	6	R33410	1.45e+00
12	125	1-5	402	19	W00238	1.45e+00
13	125	1-5	402	9	R47292	1.45e+00
14	125	1-5	402	12	R51647	1.45e+00
15	125	1-5	402	9	R50200	1.45e+00
16	125	1-5	402	9	R47252	1.45e+00
17	125	1-5	402	16	R85760	1.45e+00
18	124	1-5	402	6	R33907	1.45e+00
19	123	1-5	439	5	R28150	1.98e+00

Db 96 tpppappedrpgagnasrdgrpsggrrprprprskappkxkw 140  
QY 274 SPARPAEATSLGALSGTRHSVSGRQHAGPSTSRP-PRPW 317

RESULT 2  
ID W36855 standard; Protein; 402 AA.  
AC W36855;  
DT 10-MAR-1998 (first entry)  
DE Full length sequence of human osteogenic protein 2 (hOP-2).  
KW Human osteogenic protein; OP; OP-2; morphogen; morphogenic protein;  
KW embryogenesis; organ maintenance; tissue-specific morphogenesis;  
KW arthritis; emphysema; osteoporosis; cirrhosis.  
OS Homo sapiens.  
FH Key  
FT Region  
FT 18..263  
FT Location/Qualifiers  
FT /note= "pro region which is cleaved to yield the  
FT mature morphogenically active protein"  
Protein 264..402  
Region 301..402  
/note= "mature protein"  
Region 301..402  
/note= "conserved 7 Cys skeleton"  
US5650276-A.  
PN 22-JUL-1997. 278729.  
PF 20-JUL-1994; US-938021.  
PR 28-AUG-1992; US-938021.  
PR 30-AUG-1991; US-752764.  
PR 30-AUG-1991; US-752861.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
PI Pang RH, Rueger DC, Smart JE;  
PI WPI: 97-384665/35.  
DR N-PSDB; T97881.  
PT Screening for compounds which modulate morphogen expression - by  
PT incubating in the presence of epithelial cells which contain a  
PT cellular gene for morphogenic protein expression  
PT Disclosure: Columns 61-64; 49pp; English.  
CC The present sequence represents a human osteogenic protein-2 (hOP-2).  
CC OP-2 proteins are a group of morphogenically active proteins. Morphogens  
CC are inactive when reduced, but are active as oxidised homodimers and when  
CC oxidised with other morphogens (e.g. W36853-02). Comparison of the amino  
CC acid sequences of these morphogens has identified a consensus 6-7  
CC cysteine motif at the C-terminal. Morphogenic proteins such as OP-2 play  
CC an important role, not only in embryogenesis, but also in tissue and  
CC organ maintenance and repair in mammals. They induce a developmental  
CC cascade of tissue-specific morphogenesis in a mammal. A novel method is  
CC described for screening a candidate compound for the ability to modulate  
CC expression of a cellular gene encoding a naturally occurring morphogenic  
CC protein. The candidate compound is incubated with epithelial cells which  
CC express the cellular gene, and after a period of time the epithelial  
CC cells are assayed for the presence of or the amount of the protein  
CC expressed by the cellular gene. A change in the level of the  
CC morphogenic protein relative to the level in the epithelial cells in the  
CC absence of the candidate compound is indicative of the ability of the  
CC compound to modulate expression of the cellular gene. The method can be  
CC used to identify compounds which can increase or decrease morphogen  
CC production or levels. Such compounds can be used in the treatment of,  
CC e.g. arthritis, emphysema, osteoporosis, kidney disease, lung diseases,  
CC cardiomyopathy, and cirrhosis of the liver.  
SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 26; Length 402;  
Best Local Similarity 39.7%; Pred. No. 1.45e+00;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpr-prprappaas 67  
QY 172 GPPLYQLGAATQA-RPPPHASGRRRLGCCRANNSVREAGVPLGLPAPGARRRGGASR 230  
Db 68 rlp 70  
QY 231 SLP 233

RESULT 3  
ID R50238 standard; Protein; 402 AA.  
AC R50238;  
DT 11-OCT-1994 (first entry)  
DE Human OP-2.  
KW OP-1; OP-2; CBMP2; Vgl(fx); Vgr(fx); DPP(fx);  
KW GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);  
KW osteogenic protein; morphogen; morphogenic protein;  
KW liver; regeneration; injury; cancer; integration;  
KW transplant; gene therapy; hepatic tissue.  
OS Homo sapiens.  
PN WO9406449-A.  
PD 31-MAR-1994.  
PF 16-SEP-1993; U08808.  
PR 16-SEP-1992; US-946238.  
PR 04-MAR-1993; US-029335.  
PR 31-MAR-1993; US-040510.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
PI Pang RH, Rueger DC, Smart JE;  
PI WPI: 94-118148/14.  
DR N-PSDB; Q45164.  
PT Use of morphogen(s) to induce liver regeneration - for repair of  
PT injury, treatment of cancer etc., also to improved integration of  
PT transplant tissue, in gene therapy etc..  
PS Claim 42-42; Page 134-136; 176pp; English.  
CC Morphogens comprising an amino acid sequence sharing at least  
CC 70% homology with OP-1, OP-2, CBMP2, Vgl(fx), Vgr(fx), DPP(fx),  
CC GDF-1(fx), 60A(fx) are useful for maintaining liver function in  
CC a mammal, including means for regenerating lost or damaged hepatic  
CC tissue, means for enhancing viability and integration of hepatic  
CC tissue and organ transplants, and means for correcting liver function  
CC deficiencies, including means for enhancing diminished liver function  
CC due to tissue injury or disease.  
CC Use of BMP3(fx), BMP5(fx) and BMP6(fx) are included in the  
CC disclosure.  
SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 10; Length 402;  
Best Local Similarity 39.7%; Pred. No. 1.45e+00;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpr-prprappaas 67  
QY 172 GPPLYQLGAATQA-RPPPHASGRRRLGCCRANNSVREAGVPLGLPAPGARRRGGASR 230  
Db 68 rlp 70  
QY 231 SLP 233

RESULT 4  
ID R54937 standard; Protein; 402 AA.  
AC R54937;  
DT 15-OCT-1994 (first entry)  
DE Osteogenic protein hOP2-PP.  
KW Morphogenic protein; hOP-2-PP; OP-2; hOP2; hOP-2;  
KW tissue morphogenesis; osteogenic protein.  
OS Homo sapiens.  
PN WO9410203-A.  
PD 11-MAY-1994.  
PF 02-NOV-1993; U10520.  
PR 03-NOV-1992; US-971091.  
PR 04-MAR-1993; US-029335.  
PR 31-MAR-1993; US-040510.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
PI Pang RH, Rueger DC;  
PI WPI: 94-167392/20.  
DR N-PSDB; Q65393.  
PT A morphogenically active protein hOP-3 - for inducing tissue

```

PT morphogenesis in mammals
PS Disclosure; Page 129-131; 164pp; English.
CC A novel mouse morphogenic protein, OP3, has the sequence given in
CC R54934, and is encoded by cDNA of sequence Q65390. cDNA and protein
CC sequences were also provided for human osteogenic protein Op1
CC (Q65391, R54935), mouse Op1 (Q65392, R54936), human Op2 (Q65393,
CC R54937) and mouse Op2 (Q65394, R54938), as well as the genomic DNA
CC sequence of human Op2 (Q65395). Generic sequences given in R54939-
CC 40 accommodate homologies between Op1, Op2, Op3 and other morphogen
CC protein family members.
SQ Sequence 402 AA;

Query Match 1.58; Score 125; DB 10; Length 402;
Best Local Similarity 39.78; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 5
ID R60578 standard; Protein; 402 AA.
AC R60578;
DT 30-MAR-1995 (first entry)
DE Osteogenic protein OP2.
KW Osteogenic protein.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 239..242
FT /note="proteolytic cleavage site"
FT cleavage_site 260..263
FT /note="proteolytic cleavage site"
PN WO9420539-A.
PD 15-SEP-1994.
PF 04-MAR-1994; U02335.
PR 04-MAR-1993; US-027070.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Jones WK, Oppermann H, Ozkaynak E, Rueger DC, Sampath KT;
PI Tucker RF;
DR WPI: 94-302971/37.
DR B-PSDB; Q71426.
PT Binding partners, esp. antibodies, specific for different forms
PT of osteogenic protein - for differentiating between mature and
PT soluble complexed forms of the protein in culture media or serum.
Disclosure; Page 50-52; 70pp; English.
CC The osteogenic protein is produced recombinantly in mammalian cell
CC cultures, and may be provided to a site for bone induction in a
CC mammal with a suitable matrix to allow infiltration, proliferation
CC and differentiation of migrating progenitor cells.
SQ Sequence 402 AA;

Query Match 1.58; Score 125; DB 11; Length 402;
Best Local Similarity 39.78; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 6
ID R57973 standard; Protein; 402 AA.
AC R57973;
DT 11-OCT-1994 (first entry)

```

```

DE Human OP-2.
KW GDF-1(fx); OP-2; CBMP2; Vgl(fx); Vgr(fx); DPP(fx);
KW GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);
KW tooth socket; alveolus; osteogenic protein; morphogen;
KW morphogenic protein; periodontal tissue; regeneration;
KW tooth implant; integration; inhibition.
OS Homo sapiens.
PN WO9406399-A.
PD 31-MAR-1994.
PF 15-SEP-1993; U08742.
PR 15-SEP-1992; US-945285.
PR 04-MAR-1993; US-029335.
PR 31-MAR-1993; US-040510.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
PI Pang RHL, Rueger DC, Smart JE;
DR WPI: 94-118107/14.
DR N-PSDB; Q67313.
PT Morphogen-induced periodontal tissue regeneration - used in
PT integrating as implanted tooth in tooth socket or to inhibit
PT tissue loss associated with periodontal disease or injury
PS Claim 28-29; Page 96-98; 132pp; English.
CC Morphogens comprising an amino acid sequence sharing at least
CC 70% homology with OP-1, OP-2, CBMP2, Vgl(fx), Vgr(fx), DPP(fx),
CC GDF-1(fx), 60A(fx) and at least 80% homology with BMP3(fx),
CC BMP5(fx) and BMP6(fx) are useful for integrating an implanted
CC tooth in a tooth socket and for inhibiting tissue loss associated
CC with periodontal disease or injury.
SQ Sequence 402 AA;

Query Match 1.58; Score 125; DB 9; Length 402;
Best Local Similarity 39.78; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 7
ID R27291 standard; Protein; 402 AA.
AC R27291;
DT 26-FEB-1993 (first entry)
DE Human osteogenic protein hOP2.
KW Morphogen; morphogenic protein.
OS Homo sapiens.
PN WO9215323-A.
PD 17-SEP-1992.
PF 11-MAR-1992; U01968.
PR 11-MAR-1991; US-667274.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberampath T, Oppermann H, Pang RHL, Rueger DC;
DR WPI: 92-331475/40.
DR N-PSDB; Q28737.
PT Compans. for increasing progenitor cell population - contain a
PT morphogen to induce proliferation, useful for inhibiting
PT neoplastic growth, inducing tissue repair and in diagnosis of
PT tissue dysfunction
PS Disclosure; Page 93-95; 132pp; English.
CC Mature hOP2 is one of the preferred known morphogens which can be
CC used in the manufacture of pharmaceuticals for inducing non-
CC chondrogenic mammalian tissue growth, progenitor cell proliferation
CC and hepatic tissue growth and for maintaining the phenotypic
CC expression of differentiated cells in a mammal. Morphogenic
CC compositions of the invention can also be used to treat blood
CC disorders and impaired or lost immune function. Morphogens sharing
CC at least 70% homology with hOP2 are included.
SQ Sequence 402 AA;

```

```

Query Match          1.5%; Score 125; DB 5; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGRCERANWHSVREAGVPLGLPAPGARRRGGSASR 230

Db 68 rlp 70
QY 231 SLP 233

RESULT 8
ID R46732 standard; Protein; 402 AA.
AC R46732;
DT 25-AUG-1994 (first entry)
DE Human osteogenic pro-protein hOP2.
KW human osteogenic protein; hOP2; morphogen; infant food formulation;
KW tissue morphogenesis; tissue development; bone growth;
KW morphogen-enriched nutritional product.
OS Homo sapiens.
PN W09403075-A.
PD 17-FEB-1994.
PF 29-JUL-1993; U07190.
PR 31-JUL-1992; US-923780.
PR 31-JUL-1992; US-922813.
PR 16-SEP-1992; US-946235.
PR 04-MAR-1993; US-029335.
PR 31-MAR-1993; US-040530.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Jones WK, Kuberampath T, Oppermann H, Ozkaynak E;
PI Rueger DC, Tucker RF, Cohen CM, Pang RHL;
PI WPI: 94-065304/08.
DR N-PSDB; Q57917.
PT Morphogen enriched dietary compositions and infant formula -
PT capable of enhancing tissue morphogenesis, development and
PT viability, e.g. in infants, aged individuals and metabolic
PT disorders, e.g. anorexia nervosa, etc
PS Claim 32; Page 120-122; 160pp; English.
CC This sequence is the pro form of human osteogenic protein hOP2.
CC The mature hOP2 and proteins having at least 70% homology with it
CC are preferred morphogens for inclusion in new morphogen-enriched
CC nutritional formulations. The formulations are dietary compositions
CC suitable for people at risk for tissue damage due to protein energy
CC malnutrition or to altered metabolism function and infant
CC formulations to enhance tissue development in an infant or juvenile.
CC Sequence 402 AA;

Query Match          1.5%; Score 125; DB 9; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGRCERANWHSVREAGVPLGLPAPGARRRGGSASR 230

Db 68 rlp 70
QY 231 SLP 233

RESULT 9
ID W16366 standard; Protein; 402 AA.
AC W16366;
DT 26-AUG-1997 (first entry)
DE Human hippocampal osteogenic protein 2.
KW Human; osteogenic protein; hippocampus; soluble; ligand; antibody;
KW mature; non-covalent; dimeric; bone morphogenetic protein; purity;
KW therapeutic.
OS Homo sapiens.
PN US5610021-A.
PD 11-MAR-1997.

```

```

PF 21-FEB-1992; 841646.
PR 21-FEB-1992; US-841646.
PR 04-MAR-1993; US-027070.
PR 04-MAR-1994; US-206864.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Jones WK, Oppermann H, Ozkaynak E, Rueger DC, Sampath KT;
PI Tucker RF;
PI WPI: 97-1178399/16.
DR 17-FEB-1994.
DR N-PSDB; T73208.
PT Antibody specific for soluble form of osteogenic protein - for
PT quality control and diagnostic use
PT Disclosure; Column 25-28; 20pp; English.
PS This is the amino acid sequence of the human osteogenic protein 2 (OP2)
PS derived from hippocampal tissue. The invention relates to a novel
PS soluble form of this protein and of OPl (W16365), and especially to
PS ligands binding these proteins, e.g. poly- or monoclonal antibodies.
PS The ligands are especially able to differentiate between the soluble and
PS mature forms of these proteins. The soluble forms of the proteins
PS preferably comprises the C-terminal 6 or 7 Cys residues (i.e. for OPl
PS residues 335-431 (6 Cys) or 330-341 (7 Cys)). The soluble complex
PS comprises a pro domain of the protein non-covalently linked to a dimeric
PS form of the osteogenic protein which containing at least the soluble part
PS of OPl and another osteogenic protein e.g. OP2, bone morphogenetic
PS protein (BMP)-2, -3, -4, -5, -6 or -9. The ligand may be used to monitor
PS the purity of therapeutic osteogenic protein preparations and for
PS diagnostic purposes.
SQ Sequence 402 AA;

Query Match          1.5%; Score 125; DB 22; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGRCERANWHSVREAGVPLGLPAPGARRRGGSASR 230

Db 68 rlp 70
QY 231 SLP 233

RESULT 10
ID R44759 standard; Protein; 402 AA.
AC R44759;
DT 06-JUN-1994 (first entry)
DE Human osteogenic protein OP2.
KW Osteogenic protein; bone; cartilage; matrix; osteoarthritis;
KW repair; vascularisation; mineralisation; differentiation.
OS Homo sapiens.
PN US5266683-A.
PD 30-NOV-1993.
PF 08-APR-1988; 179406.
PR 08-APR-1988; US-179406.
PR 15-AUG-1988; US-232630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422613.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 04-DEC-1990; US-621988.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberampath T, Oppermann H, Ozkaynak E, Pang RHL;
PI Rueger DC;
PI WPI: 93-395405/49.
DR N-PSDB; Q53155.
PT New pure mammalian osteogenic proteins - induce cartilage and

```

[illegible]

PI Rueger DC, Tucker RF;  
 DR WPI; 94-065399/08.  
 DR N-PSDB; Q56233.  
 PT Use morphogens to improve survival of neural cells - also stimulating  
 PT re-differentiation in transformed cells and prodn. of adhesion  
 PT molecules, for treating traumatic injury, neuropathy and nerve cell  
 PT cancers, etc.  
 PS Claim 23; Page 131-134; 176pp; English.  
 CC This sequence represents the human hippocampus derived protein,  
 CC osteogenic protein, hop-2. The mature OP-1 protein was used in the  
 CC method of the invention for improving survival of neural cells.  
 CC Morphogens such as this, can be used to treat (protect) cells which  
 CC have suffered chemical or mechanical injury, eg. transected nerves;  
 CC demyelinated cells; cells exposed to toxins such as ethanol, and cells  
 CC at risk because of neuropathies (such as Parkinsons and Alzheimers  
 CC diseases; Huntingtons chorea; amyotrophic lateral sclerosis or  
 CC multiple sclerosis), or because of neoplastic lesions (esp. retino-  
 CC blastoma or glial cell neoplasms). At risk cells can be in the central  
 CC or peripheral nervous systems. When used to induce redifferentiation,  
 CC morphogens such as this, are used to treat neuroblastoma and then  
 CC induce formation of neurite outgrowths, cell aggregation and/or cell  
 CC adhesion. These proteins stimulate complete axonal nerve regeneration,  
 CC including vascularisation and reformation of the myelin sheath. Nerves  
 CC can be regenerated over long distances, eg. greater than 10mm.  
 CC Sequence 402 AA;  
 SQ

Query Match 1.5%; Score 125; DB 9; Length 402;  
 Best Local Similarity 39.7%; Pred. No. 1.45e+00;  
 Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvrellavlgpgr-prprappaas 67  
 QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGRCERAWNSVREAGVPLGLPAPGARRRGSASR 230  
 Db 68 rlp 70  
 QY 231 SLP 233

RESULT 14  
 ID R51647 standard; Protein; 402 AA.  
 AC R51647;  
 DT 08-JUN-1995 (first entry)  
 DE HOP2-PP prepro form human osteogenic protein 2.  
 KW HOP2-PP; prepro human osteogenic protein; osteoarthritis;  
 KW non-union fracture repair; allograft repair; orthogenesis;  
 KW cartilage and endochondrial bone formation;  
 KW periodontal, dental and craniofacial reconstruction.  
 KW Homo sapiens.  
 PN US3534557-A.  
 PD 11-OCT-1994.  
 PF 08-APR-1988; 179406.  
 PR 08-APR-1988; US-179406.  
 PR 15-AUG-1988; US-232630.  
 PR 23-FEB-1989; US-315342.  
 PR 17-OCT-1989; US-422613.  
 PR 17-OCT-1989; US-422699.  
 PR 22-FEB-1990; US-483913.  
 PR 20-AUG-1990; US-589920.  
 PR 07-SEP-1990; US-579865.  
 PR 18-OCT-1990; US-599543.  
 PR 18-OCT-1990; US-600024.  
 PA 04-DEC-1990; US-621849.  
 PA 22-FEB-1991; US-660162.  
 PA 20-DEC-1991; US-810560.  
 PA 28-JAN-1992; US-827052.  
 PA 21-FEB-1992; US-841646.  
 PA 18-DEC-1992; US-993387.  
 PI Kuberampath T, Oppermann H, Ozkaynak E, Pang RHL;  
 PI Rueger DC;  
 DR WPI; 94-324521/40.  
 DR N-PSDB; Q72706.

PT Implantable device for inducing osteogenesis - comprises porous  
 PT matrix contg. non-glycosylated dimeric, di:sulphide linked  
 PT osteogenic protein.  
 PS Claim 16; Columns 141-146; 128pp; English.  
 CC Q72706 encodes R51647 prepro human osteogenic protein 2 (hop2-pp),  
 CC fragments of this protein consisting of residues 264-402 and  
 CC 267-402, 270-402 and 243-402 are unglycosylated osteogenic polypeptides.  
 CC Any two of these polypeptides can be disulphide bonded to form a dimer,  
 CC which forms an essential component of an osteogenic protein. This  
 CC protein is dispersed in a biodegradable matrix which can be implanted  
 CC into a mammalian bone marrow cavity, here it can induce local cartilage,  
 CC bone and endochondrial bone formation; and it can also accelerate  
 CC allograft repair. This implant has the advantage of inducing all stages  
 CC of bone formation and of having a higher specific activity than other  
 CC known biosynthetic materials. The implant can be used to repair  
 CC non-union fractures and cartilage; treat osteoarthritis; and aid  
 CC in periodontal, dental or craniofacial reconstruction.  
 CC Sequence 402 AA;  
 SQ

Query Match 1.5%; Score 125; DB 12; Length 402;  
 Best Local Similarity 39.7%; Pred. No. 1.45e+00;  
 Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvrellavlgpgr-prprappaas 67  
 QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGRCERAWNSVREAGVPLGLPAPGARRRGSASR 230  
 Db 68 rlp 70  
 QY 231 SLP 233

RESULT 15  
 ID R50200 standard; Protein; 402 AA.  
 AC R50200;  
 DT 11-OCT-1994 (first entry)  
 DE Human OP-2.  
 KW OP-1; OP-2; CBMP2; Vgl(fx); Vgr(fx); DPP(fx);  
 KW GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);  
 KW osteogenic protein; morphogen; morphogenic protein;  
 KW gastrointestinal tract; luminal lining; epithelial cell;  
 KW proliferation; ulcer; lesion; inflammation; regeneration;  
 KW tissue.  
 KW Homo sapiens.  
 PN WO9406420-A.  
 PD 31-MAR-1994.  
 PF 15-SEP-1993; U088895.  
 PR 15-SEP-1992; US-945286.  
 PR 04-MAR-1993; US-029335.  
 PR 31-MAR-1993; US-040510.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Charette MF, Cohen CM, Kuberampath T, Oppermann H;  
 PI Ozkaynak E, Pang RHL, Rueger DC, Smart JE;  
 DR WPI; 94-118121/14.  
 DR N-PSDB; Q45118.  
 PT Maintaining integrity of gastrointestinal lining using a  
 PT morphogen (stimulant) - for treating or preventing ulceration,  
 PT also to inhibit endothelial cell proliferation and reduce side  
 PT effects of cancer therapy.  
 PS Claim 35-36; Page 111-113; 151pp; English.  
 CC Morphogens comprising an amino acid sequence sharing at least  
 CC 70% homology with Op-1, Op-2, CBMP2, BMP3(fx), Vgl(fx), Vgr(fx),  
 CC DPP(fx), GDF-1(fx), 60A(fx) and at least 80% homology with  
 CC BMP5(fx) and BMP6(fx) are useful for maintaining the integrity of  
 CC the gastrointestinal tract luminal lining in a mammal, including  
 CC (1) limiting epithelial cell proliferation, (2) inhibiting ulcerative  
 CC lesion formation, (3) inhibiting inflammation normally associated  
 CC with ulcerative diseases, and/or (4) stimulating the repair of  
 CC ulcerative lesions and the regeneration of the luminal tissue.  
 CC Sequence 402 AA;  
 SQ

Query Match 1.5%; Score 125; DB 9; Length 402;  
 Best Local Similarity 39.7%; Pred. No. 1.45e+00;

Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;  
 Db 12 glalcalggggpdlrppp-gc-pqrrlg-arerrdvqrellavlglpgr-prprappaas 67  
 Qy 172 GPPLYQLGAATQA-RPPPHASGPRRLGCEAWNHSYREAGVPLGLFPAGARRRGGSASR 230  
 Db 68 rlp 70  
 Qy 231 SLP 233

Search completed: Thu Jul 30 14:44:41 1998  
 Job time : 102 secs.

\*\*\*\*\*  
WILEY  
\*\*\*\*\*  
(TW)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 30 14:44:58 1998; Maspar time 42.67 Seconds  
969.188 Million cell updates/sec

Linear output not generated.

Title: >US-08-912-951-2  
Description: (1-1132) from US08912951.pep  
Perfect Score: 8465  
Sequence: 1 MPRAPRCRAVRLSLRSHYRE.....TALRAANPALPDSFKTILD 1132

Scoring table: PAM 150  
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir56  
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 53.528; Variance 118.315; scale 0.452

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	314	3.7	884	2	S3396 hypothetical protein	5.31e-30
2	162	1.9	660	1	QBEB3 BHLFI protein - human	1.01e-06
3	153	1.8	504	2	JC1306 virion protein homolo	1.61e-05
4	141	1.7	276	2	B38965 hypothetical protein	5.73e-04
5	142	1.7	310	1	PIHUSD salivary proline-rich	4.28e-04
6	136	1.6	240	2	B24264 proline-rich protein	2.43e-03
7	136	1.6	316	2	S16681 homeotic protein - hu	2.43e-03
8	132	1.6	317	2	A28996 proline-rich protein	7.54e-03
9	137	1.6	628	2	S01955 hypothetical protein	1.82e-03
10	132	1.6	924	2	S27923 gene LF3 protein - hu	7.54e-03
11	125	1.5	227	2	C29149 proline-rich protein	5.22e-02
12	123	1.5	240	2	A24264 proline-rich protein	8.97e-02
13	123	1.5	256	2	A00533 tumor-associated anti	8.97e-02
14	129	1.5	264	2	D34768 ORF4 protein - Orf vi	1.74e-02
15	131	1.5	300	2	S19560 proline-rich protein	9.98e-03
16	125	1.5	302	2	S11790 modulation protein no	5.22e-02
17	129	1.5	309	2	S10889 proline-rich protein	1.74e-02
18	125	1.5	381	2	S16506 hypothetical protein	5.22e-02
19	125	1.5	402	2	A45056 prepro osteogenic pro	5.22e-02
20	123	1.5	403	2	S52796 prpL2 protein - human	8.97e-02
21	123	1.5	439	2	S51939 chitinase (EC 3.2.1.1	8.97e-02
22	124	1.5	464	2	S23697 extensin - Volvox car	6.85e-02
23	127	1.5	566	2	S22933 testis-specific prote	3.02e-02

24 131 1.5 1106 2 JQ0405 hypothetical 119.5K p 9.98e-03  
25 117 1.4 204 2 A39066 proline-rich protein 4.39e-01  
26 115 1.4 260 2 S23773 proline-rich protein 7.36e-01  
27 116 1.4 270 2 S34361 miag protein - Salmon 5.68e-01  
28 116 1.4 295 2 B48013 proline-rich proteogl 5.68e-01  
29 118 1.4 322 1 S00054 modulation protein no 3.38e-01  
30 115 1.4 338 2 JC5707 HYA22 protein - human 7.36e-01  
31 119 1.4 347 2 S10571 epithelial tumor anti 2.60e-01  
32 121 1.4 358 1 WMBE38 infected cell protein 1.53e-01  
33 115 1.4 392 1 PIHUB6 salivary proline-rich 7.36e-01  
34 118 1.4 431 2 S09824 hypothetical protein 3.38e-01  
35 119 1.4 515 2 S10572 epithelial tumor anti 2.60e-01  
36 115 1.4 3149 1 QBEB8 BHLFI protein - human 7.36e-01  
37 114 1.3 202 2 B36795 hypothetical protein 9.50e-01  
38 114 1.3 301 2 E29149 proline-rich protein 9.50e-01  
39 114 1.3 321 1 A28663 modulation protein no 9.50e-01  
40 114 1.3 373 2 S54545 hypothetical protein 9.50e-01  
41 114 1.3 408 2 A56186 cyclin E - African cl 9.50e-01  
42 114 1.3 705 2 A35363 synapsin I splice for 9.50e-01  
43 113 1.3 1255 2 B35175 episialin B - human 1.23e+00  
44 113 1.3 1264 2 A35175 episialin A - human 1.23e+00  
45 113 1.3 1295 2 A35886 polymorphic epithelia 1.23e+00

ALIGNMENTS

RESULT 1 S53396 #type complete  
ENTRY hypothetical protein YLR318w - yeast (Saccharomyces  
TITLE cerevisiae)  
ALTERNATE\_NAMES hypothetical protein L8543.12  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change  
06-Feb-1998  
ACCESSIONS S53396  
REFERENCE S53390  
#authors Du, Z.  
#submission submitted to the EMBL Data Library, February 1995  
#description The sequence of S. cerevisiae cosmid 8543.  
#accession S53396  
#molecule\_type DNA  
#residues 1-884 #label DUZ  
#cross-references EMBL:U20618; NID:g2258165; PID:g662136; MIPS:YLR318w  
#experimental\_source strain S288C (AB972)  
GENETICS  
#gene SGD:EST2  
#map\_position 12R  
#length 884 #molecular\_weight 102662 #checksum 7604

Query Match 3.7%; Score 314; DB 2; Length 884;  
Best Local Similarity 24.3%; Pred. No. 5.31e-30;  
Matches 118; Conservative 122; Mismatches 202; Indels 44; Gaps 37;  
Db 271 LSHLSRQSPKER-VLKFFIVILQKLLPQEMFGSKKNGKIKNLLNLLSLPLNGVLPDS 329  
Qy 448 LVQLLRQSSPMQVYGFYRACLRDLRPPGLMGSRNRRERFLRNTKTKFSLGKHAQLSQE 507  
Db 330 LKLLKRLKDFWLFTS-DIWFTHKNFENLNOLAIC-FISWLFRLQIPKIQIFFYCTETS 387  
Qy 508 LTKWSVRDCAWLRSPGVCVPAEHLRREILAKFLHLMMSVYVVELLSRFFVTTET 567  
Db 388 STVTIVYF-RHDTNKLITPFIVEYFKTY-LVE-NNV-CRNHNSYTLNFNHSMRIIPK 443  
Qy 568 FQKNLRFYKSVWSKLSIGIRQHLKRVQLRELSEAEVQHQRE-ARPAALLTSRLRFIPK 626  
Db 444 KSNNEFRILAPCRGADEEFTIYKENHK-NA-IQPTOKILEYLNKRPTSF-TIYISPT 500  
Qy 627 -PDGLRPIVNDY-VVGARTFRERKRAERLTSRVKALFSVLNYERARRPGLIGASVLGID 684  
Db 501 QIADRI-KEFKORLKKFNVLPELYEMKFDVKSCYDSIPRMECHMR-ILKDAKKNGENGF 558  
Qy 685 DI-HRAWTFVLVR-RA-QDPPPELIFVKVDVTGAYDTIPQ-DRUTEVIASIKRPTONTY 740



```

JCL1306
REFERENCE
#authors Carpenter, D.E.; Misra, V.
#journal Gene (1992) 119:259-263
#title Sequences of the bovine herpesvirus 1 homologue of herpes simplex virus type-1 alpha-trans-inducing factor (UL48).
#accesion JCL1306
#molecule_type DNA
##residues 1-504 #label CAR
##cross-references EMBL:Z11610; NID:g1065725; PID:c264419; PID:g1065726
COMMENT This protein interacts with cellular transcription factors to transactivate immediate early viral genes..
GENETICS
#map_position 0.07-0.086
KEYWORDS DNA binding; transcription regulation
SUMMARY length 504 #molecular_weight 54028 #checksum 8743
Query Match 1.8%; Score 153; DB 2; Length 504;
Best Local Similarity 31.3%; Pred.No.1.61e-05;
Matches 46; Conservative 33; Mismatches 56; Indels 12; Gaps 11;
Db 351 APAEAGGWRSSGSTRTGRGAARSTTGLRQPCGGPRRAKC-CRATP-RQRLF--ARGE 406
::|:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 186 PPHASGPRLRGCRANWHSREAGVPLGLPAFCARRRGGSARSLSLPKRPARGAPE 245
::|:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 407 PRHTS-GSGAFSQ-GRPPGVRCRLGWACKARSGPARGGPGSPVRSGIGLSLR-ARGSPGP 463
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 246 PERTPVGOSWAHPGTRGPSDR-QF-CVV-S-PARPAAEATSLEGALSGTRHSHPSVGR 301
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 464 GPACGGPSRARGRRRASPNP-FGGT 489
||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 302 QHHAGPPSTRPRPDWDTPCPPIYAET 328
||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
RESULT 4 B38965 #type complete
ENTRY hypothetical protein B (insertion sequence IS1222) -
TITLE Enterobacter agglomerans
ORGANISM #Formal_name Enterobacter agglomerans
DATE 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 09-Sep-1997
ACCSSIONS B38965; S42922
REFERENCE B38965
#authors Steibl, H.D.; Lewecke, F.M.
#journal Gene (1995) 156:37-42
#title IS1222: analysis and distribution of a new insertion sequence in Enterobacter agglomerans 339.
#accesion B38965
#status Preliminary
##molecule_type DNA
##residues 1-276 #label STE
##cross-references GB:X78052; NID:g459246; PID:g459248
SUMMARY length 276 #molecular_weight 31718 #checksum 3927
Query Match 1.7%; Score 141; DB 2; Length 276;
Best Local Similarity 33.7%; Pred.No.5.73e-04;
Matches 30; Conservative 23; Mismatches 30; Indels 6; Gaps 6;
Db 42 ITELALLERR-EGYRR-TWOLLRRGLHAVNKKRYLVHLISGLVGKRER-RKGLATERL 98
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 563 VTETTQQNLRFYRKWSVKLGSIQIRHQUKRV-QLRELSEAEVQHREARPALTLTRL 621
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 99 PLL-RPAAPNLTWSDDFVDALATGRIIK 126
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 622 RFIPKPDGLRPVNMDYVVG-A-RTEREK 649
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
RESULT 5 PIHUSD #type complete
ENTRY salivary proline-rich glycoprotein precursor PRB4 (large allele)-human
TITLE basic proline-rich protein IB-5; proline-rich peptide P-D
CONTAINS #formal_name Homo sapiens #common_name man
ORGANISM 19-Feb-1984 #sequence_revision 12-Apr-1996 #text_change 19-Feb-1984
DATE
```

```

ACCESSIONS      05-Sep-1997
S03176; S03175; S10890; D25372; E38355; A03295; A61294;
REFERENCE
#authors      Lyons, K.M.; Stein, J.H.; Smithies, O.
#journal      Genetics (1988) 120:267-278
#title      Length polymorphisms in human proline-rich protein genes
              generated by intragenic unequal crossing over.
#cross-references MUID:89121440
#accession    S03176
#molecule_type DNA
#status      translation not shown
#residues    35-310 ##label LY1
#cross-references EMBL:X07715
#note      large allele
#accession    S03175
#molecule_type DNA
#status      translation not shown
#residues    35-36, 'E', 38-112, 155-310 ##label LY2
#cross-references EMBL:X07704
#note      medium allele
#accession    S10890
#molecule_type DNA
#status      preliminary; translation not shown
#residues    1-38, 60-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310
              ##label LY3
#cross-references EMBL:X07882; NID:G35647; PID:G296670
REFERENCE
#authors      Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
#journal      J. Biol. Chem. (1985) 260:11123-11130
#title      Differential RNA splicing and post-translational cleavages in
              the human salivary proline-rich protein gene system.
#cross-references MUID:85289325
#accession    D25372
#molecule_type mRNA
#residues    1-36, 'E', 38-112, 'T', 114-115, 'P', 117-121, 185-271, 'A',
              273-310 ##label MAE
REFERENCE
#authors      Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
#journal      Biochemistry (1991) 30:3351-3356
#title      Basic proline-rich proteins from human parotid saliva:
              relationships of the covalent structures of ten proteins
              from a single individual.
#cross-references MUID:91190884
#accession    E38355
#molecule_type protein
#residues    241-254, 'KN', 257-310 ##label KAU
REFERENCE
#authors      Saitoh, E.; Isemura, S.; Sanada, K.
#journal      J. Biochem. (1983) 93:495-502
#title      Complete amino acid sequence of a basic proline-rich peptide,
              P-D, from human parotid saliva.
#cross-references MUID:83186122
#accession    A03295
#molecule_type protein
#residues    241-310 ##label SAI
REFERENCE
#authors      Shimomura, H.; Kanai, Y.; Sanada, K.
#journal      J. Biochem. (1983) 93:857-863
#title      Amino acid sequences of glycopeptides obtained from basic
              proline-rich glycoprotein of human parotid saliva.
#accession    A61294
#molecule_type protein
#residues    54-57, 'E', 59-73, 'R', 82-101 ##label SHI
REFERENCE
#authors      Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.;
              McDonald, C.J.; Williamson, M.P.
#journal      FEBS Lett. (1996) 382:289-292
#title      Tannin interactions with a full-length human salivary
              proline-rich protein display a stronger affinity than with
              single proline-rich repeats.
#accession    S62891
#molecule_type protein

```

```

GENETICS
#gene      GDB:PRB4
#cross-references GDB:119514; OMIM:180990
#map_position 12p13.2-12p13.2
#introns    22/1: 34/1
#note      the list of introns may be incomplete
CLASSIFICATION
#superfamily proline-rich protein
KEYWORDS
glycoprotein; saliva; tandem repeat
FEATURE
1-16      #domain signal sequence #status predicted #label SIG\
241-310   #product proline-rich peptide P-D #status experimental
66,87,171 #label MAR\
108,150,192,213,
234       #binding_site carbohydrate (Asn) (covalent) #status
              experimental\
SUMMARY    #length 310 #molecular-weight 31351 #checksum 3960
              Predicted
Query Match 1.7%; Score 142; DB 1; Length 310;
Best Local Similarity 28.2%; Pred. No. 4.28e-04;
Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;
Db 168 QGNQSQG-PPPHPGKPERPP-OGG-NQSHRPPPP-GKPER-PPPGGQNSQG-P-PP 220
QY 177 QLGAAATQARPPPHASGPRRLGGERANVHVSREAGVPLGLPAPGARRGGASRSPLPK 236
Db 221 HPKPEGPPEQNGKSRSPGKPGQPGQEGNKPGKPGPPGKPGPPPPGPPGPPGPPQ 280
QY 237 RPRGAPEPTEPVGQGSWAHPGTRGSDR-GFCVVSAPPAE-EATSLGALSGRH 294
Db 281 APPAGKPGPPPPGPPGPPRPP 302
QY 295 SHFSVGRQHHAGPPSTRPPRP 316
RESULT 6
ENTRY      B24264 #type fragment
TITLE      proline-rich protein MP3 - mouse (fragment)
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change
              03-May-1996
ACCESSIONS B24264
REFERENCE   A92508
#authors    Ann, D.K.; Carlson, D.M.
#journal     J. Biol. Chem. (1985) 260:15863-15872
#title      The structure and organization of a proline-rich protein gene
              of a mouse multigene family.
#cross-references MUID:86059475
#accession  B24264
#molecule_type DNA
#residues    1-240 ##label ANN
CLASSIFICATION
#superfamily proline-rich protein
SUMMARY      #length 240 #checksum 5152
Query Match 1.6%; Score 136; DB 2; Length 240;
Best Local Similarity 27.2%; Pred. No. 2.43e-03;
Matches 41; Conservative 38; Mismatches 65; Indels 7; Gaps 6;
Db 24 VNGSQGPPPPGPPGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 83
QY 178 LGAAATQARPPPHASGPRRLGGERANVHVSREAGVPLGLPAPGARRGGASRSPLPKR 237
Db 84 PQGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 142
QY 238 PRGAAPPE-PTFTVGGGSWAHPG-RTRGSDRGFCVVSAPPAEATSLGALSGRH 294

```



```

#authors Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.;
#submission submitted to the EMBL Data Library, August 1990
#description Sequence and transcription of Raji Epstein-Barr virus DNA
#accession S27923
#status preliminary
#molecule_type DNA
#residues 1-924 #label PAR
#cross-references EMBL:M35547; NID:G330420; PID:G330421
SUMMARY #length 924 #molecular-weight 94304 #checksum 8997

Query Match 1.5%; Score 132; DB 2; Length 924;
Best Local Similarity 28.2%; Pred. No. 7.54e-03;
Matches 37; Conservative 33; Mismatches 52; Indels 9; Gaps 9;

Db 38 AAPRAGPPEPRTRLOPATPRRSGAADPADPVGHAA-PRAPGPEPRTRLO-PATPRRSGA 95
QY 184 ARPPPHASGPRRLGGERAWNHVSREAGVPLGLPAGARRRGGASRSLSPLPKRRRGA 243
96 ADPA-DPVGHAA-A-P-RAPGPEPRTRLOPATPRRSGAADPADPVGHAA-PRAPGPEPRTR 151
QY 244 PPERTPVQGSWAHPGRTGPDSDRG-FCVVSFARP-AEATSLGALSSTRHSH-P-SVG 300
Db 152 RLQATPRRSG 162
QY 301 RQHAGPPSTS 311

RESULT 11
ENTRY C29149 #type fragment
TITLE proline-rich protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change
20-Mar-1998
ACCESSIONS C29149
REFERENCE A92501
#authors Clements, S.; Mehansho, H.; Carlson, D.M.
#journal J. Biol. Chem. (1985) 260:13471-13477
#title Novel multigene families encoding highly repetitive peptide
sequences. Sequence analyses of rat and mouse proline-rich
protein cDNAs.
#cross-references MUID:86033799
#contents Clone PUMP40
#accession C29149
#molecule_type mRNA
#residues 1-227 #label CLE
#cross-references GB:M11902; NID:G200544; PID:G200545
CLASSIFICATION #superfamily proline-rich protein
RY #length 227 #checksum 1378

Query Match 1.5%; Score 125; DB 2; Length 227;
Best Local Similarity 27.3%; Pred. No. 5.22e-02;
Matches 41; Conservative 34; Mismatches 67; Indels 8; Gaps 8;

Db 49 QQGPPPGGPPQPPGPPGPPQPRP-PQGPPPGGPPQPPGPPGPPGPPQPRP-PQ 106
QY 177 QLGATQARPPPHAS-GPRRLGCE-RANHVSREAGVPLGLPAGARRRGGASRSLSPL 234
Db 107 GPPPPGPPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 165
QY 235 PKRPRGAAPERTPVQGSWAHPGRTGPDSDRGFCVVSFARP-AEATSLGALSSTR 293
Db 166 PRAGPPQPRPQ-GPPPTGPPQPRPQGP-PP 193
QY 294 HSPSVGRQHHAGPPSTSRPPRPWDTPCPP 323

RESULT 12
ENTRY A24264 #type fragment
TITLE proline-rich protein MP2 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change

```

```

03-May-1996
ACCESSIONS A24264
REFERENCE A92508
#authors Ann, D.K.; Carlson, D.M.
#journal J. Biol. Chem. (1985) 260:15863-15872
#title The structure and organization of a proline-rich protein gene
of a mouse multigene family.
#cross-references MUID:86059475
#accession A24264
#molecule_type DNA
#residues 1-240 #label ANN
CLASSIFICATION #superfamily proline-rich protein
SUMMARY #length 240 #checksum 5289

Query Match 1.5%; Score 123; DB 2; Length 240;
Best Local Similarity 25.8%; Pred. No. 8.97e-02;
Matches 39; Conservative 38; Mismatches 67; Indels 7; Gaps 6;

Db 24 VNGSQGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 83
QY 178 LGAATQARPPPHASGPRRLGGERAWNHVSREAGVPLGLPAGARRRGGASRSLSPLPKR 237
Db 84 PQGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 142
QY 238 PRGAPE-PERTPVQGSWAHPG-RTGPDSDRGFCVVSFARP-AEATSLGALSSTRHS 295
Db 143 PPPAGPQPRPQPPPTGPPQPRPQGP-PP 172
QY 296 HPS-VGRQHHAA--GPPSTSRPPRPWDTPCPP 323

RESULT 13
ENTRY A60533 #type complete
TITLE tumor-associated antigen DF3 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 19-Mar-1993 #sequence_revision 07-May-1993 #text_change
08-Apr-1994
ACCESSIONS A60533
REFERENCE A60533
#authors Merio, G.R.; Siddiqui, J.; Cropp, C.S.; Liscia, D.S.;
Lidereau, R.; Callahan, R.; Kufe, D.W.
#journal Cancer Res. (1989) 49:6966-6971
#title Frequent alteration of the DF3 tumor-associated antigen gene
in primary human breast carcinomas.
#accession A60533
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-256 #label MER
GENETICS
#map_position 1q21-q24
KEYWORDS glycoprotein; tandem repeat
SUMMARY #length 256 #molecular-weight 25766 #checksum 131

Query Match 1.5%; Score 123; DB 2; Length 256;
Best Local Similarity 25.3%; Pred. No. 8.97e-02;
Matches 42; Conservative 46; Mismatches 72; Indels 6; Gaps 6;

Db 35 APP-AHRVTSAPESRPAPGSTAPPAAHRTVSAPESRPAPGSTAPPAAHRTVSAPESRPAPGS 93
QY 163 APSCAVQVCGPPLYLQGAATQARPPH-ASGPRRLGGERAWNHVSREAGVPLGLPAPGA 221
Db 94 TAPPAHRTVSAPESRPAPGSTAPPAAHRTVSAPESRPAPGSTAPPAAHRTVSAPESRPAPG 152
QY 222 RRRGGSASRSLSPLPKRRRGAAPERTPVQGSWAHPGRTGPDSDRGFCVVSFARP-AE 281
Db 153 STAPPAHRTVSAPESRPAPGSTAPPAAHRTVSAPESRPAPGSTAPPV 198
QY 282 ATSLGA-LSGTRHSHFVGRQHHAGPPSTSRP-PRPW-DTPCPPV 324

RESULT 14
ENTRY D34768 #type complete
TITLE ORF4 protein - Orf virus (strain NZ2)

```



\*\*\*\*\*

WPPSRH  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run: Thu Jul 30 14:47:32 1998; MasPar time 28.16 Seconds  
1008.379 Million cell updates/sec  
Near output not generated.

Title: >US-08-912-951-2  
Description: (1-1132) from US08912951.pep  
Perfect Score: 8465  
Sequence: 1 MPRAPRCRAVRSLLRSHYRE.....TALEAANPALPSDFKTILD 1132

Scoring table: PAM 150  
Gap 11

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 55.788; Variance 96.504; scale 0.578

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query Match	Length	ID	Description	Pred. No.
1	162	1.9	650	1	YHLL_EBV	4.58e-09
2	150	1.8	234	1	PRPM_HUMAN	4.42e-07
3	153	1.8	504	1	ATIN_HSVBP	1.43e-07
4	142	1.7	276	1	PRPL_HUMAN	8.39e-06
5	133	1.6	247	1	PRP4_HUMAN	2.05e-04
6	137	1.6	628	1	V70K_TVMV	5.03e-05
7	125	1.5	261	1	PRP2_MOUSE	3.14e-03
8	131	1.5	296	1	PRP3_MOUSE	4.10e-04
9	125	1.5	302	1	NOD3_RHLP	3.14e-03
10	125	1.5	402	1	BMP8_HUMAN	3.14e-03
11	127	1.5	566	1	TS13_MOUSE	1.60e-03
12	116	1.4	270	1	MIAE_SALTY	5.84e-02
13	115	1.4	279	1	Y091_NPVP	7.99e-02
14	121	1.4	322	1	NOD1_RHSP	1.17e-02
15	115	1.4	331	1	PRP1_HUMAN	7.99e-02
16	118	1.4	399	1	BMP8_MOUSE	3.09e-02
17	118	1.4	431	1	UL61_HCMVA	3.09e-02
18	118	1.4	514	1	VE2_HPV05	3.09e-02
19	115	1.4	3149	1	TEGO_EBV	7.99e-02
20	107	1.3	174	1	BARI_CHITE	9.11e-01
21	114	1.3	202	1	VG01_HSVB	1.09e-01
22	106	1.3	233	1	YFJR_ECOLI	1.22e+00
23	106	1.3	285	1	YAFY_ECOLI	1.22e+00

24	107	1.3	318	1	NODD_RHILT	9.11e-01
25	114	1.3	321	1	NODI_BRAJA	1.09e-01
26	109	1.3	324	1	MATD_NEUCR	5.03e-01
27	114	1.3	403	1	CG1E_XENLA	1.09e-01
28	107	1.3	493	1	MFHL_MOUSE	9.11e-01
29	111	1.3	498	1	VE2_HPV08	2.75e-01
30	111	1.3	509	1	VE2_HPV36	2.75e-01
31	113	1.3	514	1	VE2_HPV5B	2.75e-01
32	107	1.3	529	1	DNB2_ADE05	1.49e-01
33	110	1.3	628	1	V70K_TVMVC	9.11e-01
34	112	1.3	633	1	LAI7_YEAST	3.72e-01
35	113	1.3	704	1	SYNI_RAT	2.03e-01
36	114	1.3	705	1	SYNI_HUMAN	1.49e-01
37	108	1.3	706	1	SYNI_BOVIN	6.78e-01
38	113	1.3	1255	1	MUC1_HUMAN	1.49e-01
39	105	1.2	286	1	AS30_HUMAN	1.63e+00
40	104	1.2	420	1	LYOX_CHICK	2.18e+00
41	104	1.2	424	1	SP49_HUMAN	2.18e+00
42	104	1.2	606	1	RP3A_MOUSE	2.18e+00
43	105	1.2	863	1	HIS2_NEUCR	1.63e+00
44	105	1.2	890	1	ACOL_ECOLI	1.63e+00
45	105	1.2	2774	1	MAPA_RAT	1.63e+00

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	660 AA.
ID	YHLL_EBV			
AC	P03181;			
DT	21-JUL-1986 (REL. 01, CREATED)			
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)			
DT	01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)			
DE	HYPOTHEICAL BHLF1 PROTEIN.			
OS	EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).			
OC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.			
RN	[1]			
RP	PROTEIN CODING REGION.			
RX	MEDLINE; 84270667.			
RA	BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J., GIBSON T.J., HATFULL G., HUDSON G.S., SATCHELL S.C., SEGUIN C., TUFFENELL P.S., BARRELL B.G.;			
RL	NATURE 310:207-211(1984).			
DR	EMBL; V01555; -; NOT_ANNOTATED_CDS.			
DR	PIR; A03742; Q0B63.			
KW	HYPOTHEICAL PROTEIN; EARLY PROTEIN; REPEAT.			
FT	DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.			
FT	REPEAT 149 273 1.			
FT	REPEAT 274 398 2.			
FT	REPEAT 399 523 3.			
FT	REPEAT 524 648 4.			
SQ	SEQUENCE 660 AA; 66244 MW; 372F08C5 CRC32;			

Query Match	1.9%;	Score 162;	DB 1;	Length 660;
Best Local Similarity	30.8%;	Pred. No. 4.58e-09;		
Matches	48;	Conservative 34;	Mismatches 60;	Indels 14; Gaps 12;
Db	251	GPPTTSRGAAG-RTURRPPGCRSARNPCPTWR--R-SCAQRGHPGAGQRPSPG 306		
QY	172	GPPLYQLGAATQARPPPHASG-PRR-R-LGGERAWHNSVREAGVPLGLPAPGA-RRRGS 227		
Db	307	TGCRPAAGCAGTTPAAGPGGGAAPSGATPHPERGSGPADPP--AAARLPPEPQBPRLP 364		
QY	228	ASLSLPLKRRRGAAPPEPR-TPVGGGNAHGPRTGRGSDRGFCVVSPARPAEATSL 286		
Db	365	QDLAAQR-C-PAGPPPTSRGA-AAQRTHR-RPPGCP 397		
QY	287	GAUSGRHSHSPSVGRGHAGPSTSRPPRPWDTFCP 322		

RESULT	2	STANDARD;	PRT;	234 AA.
ID	PRPM_HUMAN			
AC	P10161; P02813;			
DT	01-MAR-1989 (REL. 10, CREATED)			

[illegible]

```

Db 273 SVLPRTSPRRGLLPNRRHRTTSTGHIPPTTTTTSRTPGPPSRRLQRPVH-LYQSSPHTPNFRP 331
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 EG-ALSTGRHSFVSGRQHAGPPSTSRPRPMDTPCPCPVYAETKHFLL-SSGDKQLRP 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 SSTKDALLOTGPRGLHLERLGQANLRTSERSPTKRLRLPSSESNRLKPLPEATLAP 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 SFLL-SSL-R-PSLTGARELVTETIFLGRSPWMPGTFRRLPLQRYWQMRPLFLLG- 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 SYRHRREYVLLPNPPAALPSIAVTSSRGKTIHSLPKGALPK-EGAPPPPRRL 442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 399 NHAQC-PYGVLTKTHCPRLA-AVTPAGVCAREKPGSGVAAPPEEDTPRRL 448
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID PRP2_MOUSE STANDARD; PRT; 261 AA.
AC P05142;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86059475.
RA ANN D.K., CARLSON D.M.;
RL J. BIOL. CHEM. 260:15863-15872(1985).
DR EMBL; M12099; G200547; -.
DR HSSP; P19999; 1CLG.
KW REPEAT; SALIVA; SIGNAL.
FT SIGNAL 1 15
FT CHAIN 16 261
SQ SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;
      POTENTIAL.
      PROLINE-RICH PROTEIN MP-2.

Query Match 1.5%; Score 125; DB 1; Length 261;
Best Local Similarity 27.0%; Pred. No. 3.14e-03;
Matches 40; Conservative 33; Mismatches 69; Indels 6; Gaps 5;

Db 76 PPOGPPPGGQPRPPGQPPPPGQPPRPPGQPPPGGQPPGQPPGQPPGQPPGQPPGQPPGQ 135
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 ATQARPPPHASGPRRLGCRRAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRR 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 PPPGGQQLPPOGPPPPAGP-QPRPPGQPPPPAGQPPRPPGQPPPTTGQPRPTGQPPPT 194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 GAAP-PE-RTPYQGVGSWAHPGTRGSDRGFCVSPARPAEATLEGALSGTRHSF 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 GGPQQPPGQPPPPGQPPRPPGQPPPP 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 VGRQHHN--GPPSTSRP-PRPNDTPCPC 323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID PRP3_MOUSE STANDARD; PRT; 296 AA.
AC P05143;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLINE-RICH PROTEIN MP-3 (FRAGMENT).
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86059475.
RA ANN D.K., CARLSON D.M.;
RL J. BIOL. CHEM. 260:15863-15872(1985).
DR EMBL; M12100; G200549; -.
DR HSSP; P19999; 1CLG.
KW REPEAT; SALIVA.
FT NON TER 1

```



SQ SEQUENCE 296 AA; 29521 MW; C61DDCC6 CRC32;

Query Match 1.5%; Score 131; DB 1; Length 296;  
Best Local Similarity 26.4%; Pred. No. 4.10e-04;  
Matches 39; Conservative 38; Mismatches 64; Indels 7; Gaps 6;

Db 16 SGSPRPVNSQQPPPGGQPPRP-PQGPPPGQPPRP-PQGPPPGQPPRP-PQGP 73  
QY 180 AATQARPPPHAS--GPRRLGCE-RAWNSVREAGVPLGLPAPGARRGGGSASRLPLPK 236  
Db 74 PPGGPPRPQPPPGGQPPRP-PQGPQPPPGQPPPGQPPPGQPPPGQPPPGQPP 132  
QY 237 RPRGAPEPPTVPGQSWAHPTGTRGSDRGFCVSPAR-PAEATSLGALSGTRHS 295  
Db 133 GGQPPRPQPPPGQPPPGQPPPGQPPPGQPPPGQPPPGQPPPGQPPPGQPP 160  
QY 296 HPSVGRQHAGPPTSTSRPPRPWDTCPP 323

RESULT 9  
NOD3\_RHILP STANDARD; PRT; 302 AA.  
P23720;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE NODULATION PROTEIN D III.  
GN NODD3.  
OS RHIZOBIUM LEGUMINOSARUM (BIOVAR PHASEOLI).  
OG PLASMID SYM.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC RHIZOBIACEAE.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-8002;  
RX MEDLINE: 91014692.  
RA DAVIS E.O., JOHNSTON A.W.B.;  
RL MOL. MICROBIOL. 4:921-932(1990).  
CC -1- FUNCTION: NODD REGULATES THE EXPRESSION OF THE NOD ABCFE GENES  
WHICH ENCODE OTHER NODULATION PROTEINS. NODD IS ALSO A NEGATIVE  
REGULATOR OF ITS OWN EXPRESSION. BINDS FLAVONOIDS AS INDUCERS.  
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL: X54215; G46237; -.  
DR PIR: S11790; S11790.  
DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
KW NODULATION; TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR;  
KW REPRESSOR; PLASMID.  
SQ SEQUENCE 302 AA; 34274 MW; ED62D2FA CRC32;

Query Match 1.5%; Score 125; DB 1; Length 302;  
Best Local Similarity 33.0%; Pred. No. 3.14e-03;  
Matches 35; Conservative 23; Mismatches 40; Indels 8; Gaps 8;

Db 32 LSQPAMSAVRL-RSVFRDEL-FTMRGEFVTP-R-AEDLAPAIRALQHRLNIIPW 87  
QY 1 MPRAPRCRAVRSLLSHRYEVLPLATEFVRLGPGQWRLVQGDPAFRALVAQCLVCPW 60  
Db 88 DKFTPDQSDRIFR-VS-LCDFVTVVLFQKILERLAREAPGISFDLL 131  
QY 61 DA-RPPPAAPSFQVSCLEKELVARVL-QLRCERGAKNVLAFGFALL 104

RESULT 10  
BMP8\_HUMAN STANDARD; PRT; 402 AA.  
AC P34820;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE BONE MORPHOGENETIC PROTEIN 8 PRECURSOR (BMP-8) (OSTEOGENIC PROTEIN 2) (OP-2).  
GN BMP8.  
OS HOMO SAPIENS (HUMAN).

Query Match 1.5%; Score 125; DB 1; Length 402;  
Best Local Similarity 39.7%; Pred. No. 3.14e-03;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 GLALCALGGGGLRPPP-GC-PQRLG-ARERRDVQREILAVGLDGR-PPRPAPPAAS 67  
QY 172 GPPLYQLGAATQA-RPPPHASPRRLGCRANHSVREAGVPLGLPAPGARRGGGSASR 230  
Db 68 RLP 70  
QY 231 SLP 233

Query Match 1.5%; Score 125; DB 1; Length 402;  
Best Local Similarity 39.7%; Pred. No. 3.14e-03;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 GLALCALGGGGLRPPP-GC-PQRLG-ARERRDVQREILAVGLDGR-PPRPAPPAAS 67  
QY 172 GPPLYQLGAATQA-RPPPHASPRRLGCRANHSVREAGVPLGLPAPGARRGGGSASR 230  
Db 68 RLP 70  
QY 231 SLP 233

RESULT 11  
TS13\_MOUSE STANDARD; PRT; 566 AA.  
AC Q01755;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TESTIS-SPECIFIC PROTEIN PBS13.  
GN TCPI1.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-CBA/CA; TISSUE-TESTIS;  
RX MEDLINE: 91372153.  
RA MAZARAKIS N.D., NELKI D., LYON M.F., EVANS E.P., RUDDY S.,  
RA FREEMONT P., DUDLEY K.;  
RL DEVELOPMENT 111:561-571(1991).  
CC -1- FUNCTION: POSSIBLY PLAYS AN IMPORTANT ROLE IN SPERM DEVELOPMENT  
AND FUNCTION.  
CC -1- TISSUE SPECIFICITY: TESTIS.  
CC -1- DEVELOPMENTAL STAGE: FIRSTLY EXPRESSED IN THE PACHYTENE  
SPERMATOCYTE STAGE.  
CC -1- SIMILARITY: SOME TO YEAST SOK1.  
DR EMBL: X52128; G54853; -.  
DR PIR: S22933; S22933.  
DR MGD: MGI:98544; TCPI1.  
KW TESTIS; SPERMATOGENESIS; REPEAT; COILED COIL.  
FT DOMAIN 171 206  
COILED COIL (5 HEPTADS).

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=HIPPOCAMPUS;  
RX MEDLINE: 93094231.  
RA OZKAYNAK E., SCHNEGELSBERG P.N.J., JIN D.F., CLIFFORD G.M.,  
RA WARREN F.D., DRIER E.A., OPPERMAN H.;  
RL J. BIOL. CHEM. 267:25220-25227(1992).  
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE  
OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF  
EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION  
AND BONE HOMEOSTASIS (BY SIMILARITY).  
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.  
CC -1- SIMILARITY: TO OTHER GROWTH FACTORS OF THE TGF-BETA FAMILY.  
DR EMBL: M97016; G189390; -.  
DR PIR: A45056; A45056.  
DR HSP: P08112; 11FG.  
DR PROSITE: PS00250; TGF\_BETA; 1.  
KW SIGNAL; GROWTH FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.  
FT SIGNAL 1 19  
FT PROPEP 20 263  
FT CHAIN 264 402  
FT DISULFID 301 367  
FT DISULFID 330 399  
FT DISULFID 334 401  
FT DISULFID 366 366  
FT CARBOHYD 158 158  
FT CARBOHYD 343 343  
SQ SEQUENCE 402 AA; 44764 MW; 8F20C61B CRC32;

[illegible]



\*\*\*\*\*

WQSRLEH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Jul 30 14:49:07 1998; MasPar time 50.56 Seconds  
942.752 Million cell updates/sec  
Near output not generated.

Title: >US-08-912-951-2  
Description: (J-1132) from US08912951.pep  
Perfect Score: 8465  
Sequence: 1 MPRAPRCRAVRLSLRSHYRE.....TALEAANPALPSPDFKTILD 1132

Scoring table: PAM 150  
Gap 11

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl5  
1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 53.898; Variance 101.765; scale 0.530

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	8465	100.0	1132	2	TELOMERASE REVERSE TRA	0.00e+00
2	8459	99.9	1132	2	TELOMERASE CATALYTIC S	0.00e+00
3	678	8.0	988	1	TELOMERASE REVERSE TRA	6.18e-110
4	681	8.0	989	1	TELOMERASE REVERSE TRA	1.42e-110
5	439	5.2	1031	3	TELOMERASE SUBUNIT P12	3.85e-60
6	424	5.0	67	10	TELOMERASE CATALYTIC S	4.19e-57
7	314	3.7	884	1	CHROMOSOME XII COSMID	2.25e-35
8	142	1.7	234	2	PAROTID 'O' PROTEIN (F	3.44e-05
9	141	1.7	276	9	IS 1222 GENE ORF-A AND	4.83e-05
10	147	1.7	286	11	HYPOTHETICAL PROTEIN (	6.21e-06
11	143	1.7	574	13	ACTIN ASSOCIATED PROTE	2.45e-05
12	136	1.6	316	2	HOMEOBOX PROTEIN (FRAG	2.58e-04
13	132	1.6	317	10	PROLINE RICH PROTEIN P	9.59e-04
14	133	1.6	539	3	SIMILARITY TO COLLAGEN	6.92e-04
15	135	1.6	585	11	HYPOTHETICAL 60.2 KD P	3.59e-04
16	132	1.6	924	11	LF3 PROTEIN.	9.59e-04
17	125	1.5	227	10	PROLINE-RICH SALIVARY	9.02e-03
18	129	1.5	264	11	HOMOLOGUE OF RETROVIRA	2.53e-03
19	131	1.5	300	10	PROLINE RICH PROTEIN.	1.33e-03
20	129	1.5	309	2	SALIVARY PROLINE-RICH	2.53e-03

21	123	1.5	403	2	Q15220	PREL-2 PROTEIN (FRAGME	1.69e-02
22	123	1.5	439	8	Q42421	CHITINASE PRECURSOR.	1.69e-02
23	124	1.5	464	8	Q41645	EXTENSIN (FRAGMENT).	1.23e-02
24	125	1.5	501	2	Q99958	MESENCHYME FORK HEAD-1	9.02e-03
25	128	1.5	527	9	Q53913	BETA-KETOACYL SYNTHASE	3.48e-03
26	130	1.5	1236	9	Q06264	HYPOTHETICAL 131.9 KD	1.83e-03
27	126	1.5	3247	11	Q65533	UL36.	6.58e-03
28	115	1.4	106	11	Q41981	HYPOTHETICAL 9.9 KD PR	1.92e-01
29	117	1.4	197	3	Q17626	C04G2.8.	1.06e-01
30	118	1.4	203	3	Q91497	CODED FOR BY C. ELEGAN	7.82e-02
31	115	1.4	203	3	Q91026	CODED FOR BY C. ELEGAN	1.92e-01
32	118	1.4	236	11	Q85028	SUBSTRATE OF THE PROTE	7.82e-02
33	115	1.4	260	10	Q64306	PROLINE-RICH PROTEIN.	1.92e-01
34	116	1.4	295	10	Q07611	PROLINE-RICH PROTEOGLY	1.43e-01
35	115	1.4	297	2	Q16038	PRB1M PROTEIN (FRAGMEN	1.92e-01
36	118	1.4	322	9	Q53180	MPIK3030 NODD1 GENE IN	7.82e-02
37	115	1.4	340	2	Q15194	HYA22.	1.92e-01
38	115	1.4	358	2	Q08805	SALIVARY PROLINE-RICH	1.92e-01
39	115	1.4	382	2	Q00599	CON1.	1.92e-01
40	116	1.4	488	8	Q04528	SEQUENCE OF BAC F20P5	1.43e-01
41	118	1.4	526	11	Q85027	VIRAL PROTEINASE.	7.82e-02
42	122	1.4	657	8	Q22835	SF16 ISOLOG.	2.30e-02
43	114	1.3	185	11	Q39738	ORF 1 PROTEIN.	2.58e-01
44	114	1.3	301	10	Q62105	31-KDA PROLINE-RICH SA	2.58e-01
45	114	1.3	373	1	Q03787	HYPOTHETICAL 43.8 KD P	2.58e-01

## ALIGNMENTS

RESULT 1  
ID Q14746 PRELIMINARY; PRT; 1132 AA.  
AC Q14746;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE TELOMERASE REVERSE TRANSCRIPTASE.  
GN HTRT.

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,  
RA LINGNER J., HARLEY C.B., CECCH T.R.;  
RL SCIENCE 277:955-959(1997).  
DR EMBL; AF015950; G2330017; -.  
KW RNA-DIRECTED DNA POLYMERASE.  
SQ SEQUENCE 1132 AA; 126996 MW; 2DFBEDF3 CRC32;

Query Match 100.0%; Score 8465; DB 2; Length 1132;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	MPRAPRCRAVRLSLRSHYREVLP	PLATEVRLPGQWGLVQGDPAARALVAQCLVCVPW	60
QY	1	MPRAPRCRAVRLSLRSHYREVLP	PLATEVRLPGQWGLVQGDPAARALVAQCLVCVPW	60
Db	61	DARPPPAAPSFROVSCILKELVARVQLRCERGA	KNVLAFFALLDARGGPPFAFTTSVR	120
QY	61	DARPPPAAPSFROVSCILKELVARVQLRCERGA	KNVLAFFALLDARGGPPFAFTTSVR	120
Db	121	SYLPNTVTDALRGSGANGLLRLRRVGGDVLVHLL	ARCAFLVLPASCAVQCGPPLYQLGA	180
QY	121	SYLPNTVTDALRGSGANGLLRLRRVGGDVLVHLL	ARCAFLVLPASCAVQCGPPLYQLGA	180
Db	181	ATQARPPPHASGPRRLGCGERANWHSVREAGVPL	GLPAPGARRGGGSASRLPLPKPRR	240
QY	181	ATQARPPPHASGPRRLGCGERANWHSVREAGVPL	GLPAPGARRGGGSASRLPLPKPRR	240
Db	241	GAAPPEPTVPGGQSWAHPTGTRGSDRGFCVSPAR	PAEATSLLEGALSGTRSHHSVG	300
QY	241	GAAPPEPTVPGGQSWAHPTGTRGSDRGFCVSPAR	PAEATSLLEGALSGTRSHHSVG	300

Db 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGDKEQLRPSFLSSLRPSLTGARRL 360  
QY 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGDKEQLRPSFLSSLRPSLTGARRL 360  
Db 361 VETIFGSRPMPGTPRLPRLPORYWQMPRLFLELGNHAQCPYGVLLKTHCPRLRAVT 420  
QY 361 VETIFGSRPMPGTPRLPRLPORYWQMPRLFLELGNHAQCPYGVLLKTHCPRLRAVT 420  
Db 421 PAAGVCAREKPOGSAAPAEEDTDPRLLVOLLROHSSPQWYGVFVRACLRLVPPGLWGS 480  
QY 421 PAAGVCAREKPOGSAAPAEEDTDPRLLVOLLROHSSPQWYGVFVRACLRLVPPGLWGS 480  
Db 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540  
QY 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540  
Db 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 600  
QY 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 600  
Db 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPVNMDDYVVGARTFREKRAERLTSRVKA 660  
QY 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPVNMDDYVVGARTFREKRAERLTSRVKA 660  
Db 661 LFSVLNTERARRPGLLGASVGLDLDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTI 720  
QY 661 LFSVLNTERARRPGLLGASVGLDLDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTI 720  
Db 721 PQDLRLTEVIAIIPQNTYCVRRYAVVQKAAHGVRAFKSHVSTLTLQPYMRQFVAHL 780  
QY 721 PQDLRLTEVIAIIPQNTYCVRRYAVVQKAAHGVRAFKSHVSTLTLQPYMRQFVAHL 780  
Db 781 QETSPLRDVAVVIEQSSSLNEASSGLFVFLRMCCHAVRIRGKSVOCQIGIPQGSILSTL 840  
QY 781 QETSPLRDVAVVIEQSSSLNEASSGLFVFLRMCCHAVRIRGKSVOCQIGIPQGSILSTL 840  
Db 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 1020  
QY 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 1020  
Db 1021 FHOQVKNPTFFLRVISTDASLCYSILKAKNAGSLGAKGAAGPLPSEAVQWLCHQAFLL 1080  
QY 1021 FHOQVKNPTFFLRVISTDASLCYSILKAKNAGSLGAKGAAGPLPSEAVQWLCHQAFLL 1080  
Db 1081 KLTRHRVTVPLLGSLRTAQQLSRKLPGLTTLTALEAANPALPSDFKTILD 1132  
QY 1081 KLTRHRVTVPLLGSLRTAQQLSRKLPGLTTLTALEAANPALPSDFKTILD 1132

## RESULT 2

ID 014783 PRELIMINARY; PRT; 1132 AA.  
AC 014783;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE TELOMERASE CATALYTIC SUBUNIT.  
GN HEST2.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEYERSON M., COUNTER C.M., EATON E.N., ELLISEN L.W., STEINER P.,  
RA CADDLE S.D., ZIAUGRA L., BEIJERSBERGEN R.L., DAVIDOFF M.J., LIU Q.,

RA BACCHETTI S., HABER D.A., WEINBERG R.A.;  
RL CELL 90:785-795(1997).  
DR EMBL; AF018167; G2347129; -.  
SQ SEQUENCE 1132 AA; 126938 MW; C1E5E2AF CRC32;

Query Match 99.98; Score 8459; DB 2; Length 1132;  
Best Local Similarity 99.98; Pred. No. 0.00e+00;  
Matches 1131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MPRAPRCRAVRSLLRSHYREVLPPLATEFVRRLPGQWRLVQGDPAAPRALVAQCLVCPW 60  
QY 1 MPRAPRCRAVRSLLRSHYREVLPPLATEFVRRLPGQWRLVQGDPAAPRALVAQCLVCPW 60  
Db 61 DARPPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFAFALLDARGGPPPEAFTSVR 120  
QY 61 DARPPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFAFALLDARGGPPPEAFTSVR 120  
Db 121 SYLPTNTVDALRGSGANGLLLRVGGDDVHLARCAFLVLAAPSCAYQVCGPPLYQLGA 180  
QY 121 SYLPTNTVDALRGSGANGLLLRVGGDDVHLARCAFLVLAAPSCAYQVCGPPLYQLGA 180  
Db 181 ATQARPPPHASGPPRRRLGCERAWNHSVREAGVPLGLPAGARRRGGGSASRLPLPKPRR 240  
QY 181 ATQARPPPHASGPPRRRLGCERAWNHSVREAGVPLGLPAGARRRGGGSASRLPLPKPRR 240  
Db 241 GAAPEPERTVPGOGSWAHPGTRGSDRGFCVSPARPAAEEATSLEGALSGTRHSHPSVG 300  
QY 241 GAAPEPERTVPGOGSWAHPGTRGSDRGFCVSPARPAAEEATSLEGALSGTRHSHPSVG 300  
Db 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGDKEQLRPSFLSSLRPSLTGARRL 360  
QY 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGDKEQLRPSFLSSLRPSLTGARRL 360  
Db 361 VETIFGSRPMPGTPRLPRLPORYWQMPRLFLELGNHAQCPYGVLLKTHCPRLRAVT 420  
QY 361 VETIFGSRPMPGTPRLPRLPORYWQMPRLFLELGNHAQCPYGVLLKTHCPRLRAVT 420  
Db 421 PAAGVCAREKPOGSAAPAEEDTDPRLLVOLLROHSSPQWYGVFVRACLRLVPPGLWGS 480  
QY 421 PAAGVCAREKPOGSAAPAEEDTDPRLLVOLLROHSSPQWYGVFVRACLRLVPPGLWGS 480  
Db 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540  
QY 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540  
Db 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 600  
QY 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 600  
Db 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPVNMDDYVVGARTFREKRAERLTSRVKA 660  
QY 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPVNMDDYVVGARTFREKRAERLTSRVKA 660  
Db 661 LFSVLNTERARRPGLLGASVGLDLDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTI 720  
QY 661 LFSVLNTERARRPGLLGASVGLDLDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTI 720  
Db 721 PQDLRLTEVIAIIPQNTYCVRRYAVVQKAAHGVRAFKSHVSTLTLQPYMRQFVAHL 780  
QY 721 PQDLRLTEVIAIIPQNTYCVRRYAVVQKAAHGVRAFKSHVSTLTLQPYMRQFVAHL 780  
Db 781 QETSPLRDVAVVIEQSSSLNEASSGLFVFLRMCCHAVRIRGKSVOCQIGIPQGSILSTL 840  
QY 781 QETSPLRDVAVVIEQSSSLNEASSGLFVFLRMCCHAVRIRGKSVOCQIGIPQGSILSTL 840  
Db 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 900  
QY 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRLLFFYRKSVWSKLQSIGIRQHLKRVOLRE 900  
Db 901 RKTVVNFPVEDEALGGTAFTVQMPAHGLFPWCGLLDTLRTLEVOGSDYSYARTSTRASLT 960  
QY 901 RKTVVNFPVEDEALGGTAFTVQMPAHGLFPWCGLLDTLRTLEVOGSDYSYARTSTRASLT 960

Db 961 NRGFAGNRMRKLVGLVRLKCHSLDLOVNSLOVCTNYIKILLLOAYRFHACVLOLP 1020  
 QY 961 NRGFAGNRMRKLVGLVRLKCHSLDLOVNSLOVCTNYIKILLLOAYRFHACVLOLP 1020  
 Db 1021 FHOQVKNFTFFLRVSDTASLCYSILKAKNAGMSLGAKAAGPLPSEAYOWLCHQAFLL 1080  
 QY 1021 FHOQVKNFTFFLRVSDTASLCYSILKAKNAGMSLGAKAAGPLPSEAYOWLCHQAFLL 1080  
 Db 1081 KLTHRVTVVPLGLSRLTAQTOISRLPLGTTLTALAAAAANPALPSPDKTILD 1132  
 QY 1081 KLTHRVTVVPLGLSRLTAQTOISRLPLGTTLTALAAAAANPALPSPDKTILD 1132

## RESULT 3

ID O13339 PRELIMINARY; PRT; 988 AA.  
 AC O13339;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.  
 GN TRT1.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972H-;  
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,  
 RA LINGNER J., HARLEY C.B., CECCH T.R.;  
 RL SCIENCE 277:955-959(1997).  
 DR EMBL; AF015783; G2340168; -.  
 KW RNA-DIRECTED DNA POLYMERASE.  
 SQ SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;

Query Match 8.0%; Score 678; DB 1; Length 988;  
 Best Local Similarity 26.4%; Pred. No. 6.18e-110;  
 Matches 129; Conservative 133; Mismatches 208; Indels 18; Gaps 15;  
 Db 340 PNOVFAFLRSILVRFPKLIWGNORIFEILKOLETFLKLSRYESFSLHYLMSNIKISEI 399  
 QY 458 PQOYGFVRACLRLVPPGLWGRSHNRERFLRNTKFKISLGKHAQLSLOELTWKMSYRDC 517  
 Db 400 EWLVLGRKSNAMKCLSDFEKR-KQ-IPAEFYIWLNSFIPILOSFFYITESSDLNRNTV 457  
 QY 518 AWWL---RRSPGVGCVPAAEHRLREELILAKFLHLMVYVVELLSRFFVYVTTTFQKNRLF 574  
 Db 458 YFRKDIW-KLLCRPFITSMKMEAFKINENNVRMDTQ-KTTLPPAVIRLLPKKNTFRLIT 515  
 QY 575 FYRKSVMKLSQSIGIRHQLRVQLRELSEAEVROHREARPALTSRLRFPKPDGLRPV 634  
 Db 516 NLKRFELIKMGSNKKMLVSTNQTLPVASTIKHLINSESSGIPFN-LEVYMKLLTFKKDL 574  
 QY 635 NMDYVVGARTFREKRAERLTSVKALFSLVLYERARRPGLLGASVGLDDIHRWRTFV 694  
 Db 575 LKRMFGRK-K-YFVRIDIKSCYDRKQDLMFRIYKVKKLKDP-EVIRKYATIH-ATS 630  
 QY 695 LRVRADPPPELYFVKVDVTGAYDTIPQDRLETVIASIIPQNTYCVRRYAVVQKAAHG 754  
 Db 631 ATKNFVSEAFSFDVMPFEK--VVOLLSMT--SDTLFVDVFDYTWKSSSEIFKMLKEHLS 687  
 QY 755 VRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDVAVIEQSSSLNEASSGLFDVFLRMC 814  
 Db 688 GHVIVKIGNSQYLQKVGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFLFI 747  
 QY 815 HIAVRIRGKSYVQCGIPQGSILSLCSLCYGDMEKFLFAGIRRDG-LLLRLVDDFL 873  
 Db 748 TVNKKDAKFLNLSRGFEKHNFTSLEKTVINFNENSGIINNTFFNESKRR-M-PFFGF 805  
 QY 874 TPLHTHAKTFLRLTVRGVPEYGCYVNLKRTVYVNFPEDEALGCTAFVQMPAHGLFPWCG 933  
 Db 806 SVNMRSLD 813  
 QY 934 LLDTRTLE 941

Query Match 8.0%; Score 678; DB 1; Length 988;  
 Best Local Similarity 26.4%; Pred. No. 6.18e-110;  
 Matches 129; Conservative 133; Mismatches 208; Indels 18; Gaps 15;

Db 340 PNOVFAFLRSILVRFPKLIWGNORIFEILKOLETFLKLSRYESFSLHYLMSNIKISEI 399  
 QY 458 PQOYGFVRACLRLVPPGLWGRSHNRERFLRNTKFKISLGKHAQLSLOELTWKMSYRDC 517  
 Db 400 EWLVLGRKSNAMKCLSDFEKR-KQ-IPAEFYIWLNSFIPILOSFFYITESSDLNRNTV 457  
 QY 518 AWWL---RRSPGVGCVPAAEHRLREELILAKFLHLMVYVVELLSRFFVYVTTTFQKNRLF 574  
 Db 458 YFRKDIW-KLLCRPFITSMKMEAFKINENNVRMDTQ-KTTLPPAVIRLLPKKNTFRLIT 515  
 QY 575 FYRKSVMKLSQSIGIRHQLRVQLRELSEAEVROHREARPALTSRLRFPKPDGLRPV 634  
 Db 516 NLKRFELIKMGSNKKMLVSTNQTLPVASTIKHLINSESSGIPFN-LEVYMKLLTFKKDL 574  
 QY 635 NMDYVVGARTFREKRAERLTSVKALFSLVLYERARRPGLLGASVGLDDIHRWRTFV 694  
 Db 575 LKRMFGRK-K-YFVRIDIKSCYDRKQDLMFRIYKVKKLKDP-EVIRKYATIH-ATS 630  
 QY 695 LRVRADPPPELYFVKVDVTGAYDTIPQDRLETVIASIIPQNTYCVRRYAVVQKAAHG 754  
 Db 631 ATKNFVSEAFSFDVMPFEK--VVOLLSMT--SDTLFVDVFDYTWKSSSEIFKMLKEHLS 687  
 QY 755 VRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDVAVIEQSSSLNEASSGLFDVFLRMC 814  
 Db 688 GHVIVKIGNSQYLQKVGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFLFI 747  
 QY 815 HIAVRIRGKSYVQCGIPQGSILSLCSLCYGDMEKFLFAGIRRDG-LLLRLVDDFL 873  
 Db 748 TVNKKDAKFLNLSRGFEKHNFTSLEKTVINFNENSGIINNTFFNESKRR-M-PFFGF 805  
 QY 874 TPLHTHAKTFLRLTVRGVPEYGCYVNLKRTVYVNFPEDEALGCTAFVQMPAHGLFPWCG 933  
 Db 806 SVNMRSLD 813  
 QY 934 LLDTRTLE 941

## RESULT 4

ID O13338 PRELIMINARY; PRT; 989 AA.  
 AC O13338;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.  
 GN TRT1.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972H-;  
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,  
 RA LINGNER J., HARLEY C.B., CECCH T.R.;  
 RL SCIENCE 277:955-959(1997).  
 DR EMBL; AF015783; G2340169; -.  
 KW RNA-DIRECTED DNA POLYMERASE.  
 SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;

Query Match 8.0%; Score 681; DB 1; Length 989;  
 Best Local Similarity 26.3%; Pred. No. 1.42e-110;  
 Matches 129; Conservative 136; Mismatches 204; Indels 21; Gaps 18;

Db 340 PNOVFAFLRSILVRFPKLIWGNORIFEILKOLETFLKLSRYESFSLHYLMSNIKISEI 399  
 QY 458 PQOYGFVRACLRLVPPGLWGRSHNRERFLRNTKFKISLGKHAQLSLOELTWKMSYRDC 517  
 Db 400 EWLVLGRKSNAMKCLSDFEKR-KQ-IPAEFYIWLNSFIPILOSFFYITESSDLNRNTV 457  
 QY 518 AWWL---RRSPGVGCVPAAEHRLREELILAKFLHLMVYVVELLSRFFVYVTTTFQKNRLF 574  
 Db 458 YFRKDIW-KLLCRPFITSMKMEAFKINENNVRMDTQ-KTTLPPAVIRLLPKKNTFRLIT 515  
 QY 575 FYRKSVMKLSQSIGIRHQLRVQLRELSEAEVROHREARPALTSRLRFPKPDGLRPV 634  
 Db 516 NLKRFELIKMGSNKKMLVSTNQTLPVASTIKHLINSESSGIPFN-LEVYMKLLTFKK 573  
 QY 635 NMDYVVGARTFREKRAERLTSVKALFSLVLYERARRPGLLGASVGLDDIHRWRTFV 692  
 Db 574 LKRMFGRK-K-YFVRIDIKSCYDRKQDLMFRIYKVKKLKDP-EVIRKYATIH-ATS 629  
 QY 693 FLVRADPPPELYFVKVDVTGAYDTIPQDRLETVIASIIPQNTYCVRRYAVVQKAAH 752  
 Db 630 DRATKNFVSEAFSFDVMPFEK--VVOLLSMT--SDTLFVDVFDYTWKSSSEIFKMLKEH 686  
 QY 753 GHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDVAVIEQSSSLNEASSGLFDVFLR 812  
 Db 687 LSHGIVKIGNSQYLQKVGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFL 746  
 QY 813 MCHAVIRGKSYVQCGIPQGSILSLCSLCYGDMEKFLFAGIRRDG-LLLRLVDDFL 871  
 Db 747 FITVKKDAKFLNLSRGFEKHNFTSLEKTVINFNENSGIINNTFFNESKRR-M-PFF 804  
 QY 872 LVPHLTHAKTFLRLTVRGVPEYGCYVNLKRTVYVNFPEDEALGCTAFVQMPAHGLFPWC 931  
 Db 805 GFSVNMRLSD 814  
 QY 932 GLLDTRTLE 941

## RESULT 5

ID O00939 PRELIMINARY; PRT; 1031 AA.  
 AC O00939;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE TELOMERASE SUBUNIT P123.  
 OS EUKARYOTA; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA;  
 OC EUKARYOTA; EUPLOTIDA; EUPLOTES.  
 RN [1]

```
RP SEQUENCE FROM N.A.
RX MEDLINE; 97274210.
RA LINGNER J., HUGHES T.R., SHEVCHENKO A., MANN M., LUNDBLAD V.,
RA CECH T.R.;
RL SCIENCE 276:561-567(1997).
DR EMBL; U95964; G2072336; -.
SQ SEQUENCE 1031 AA; 122562 MW; 21A885CD CRC32;

Query Match
Best Local Similarity 5.2%; Score 439; DB 3; Length 1031;
Matches 142; Conservative 168; Mismatches 261; Indels 47; Gaps 34;

Db 361 FINEFFNIILPKDELTCR-NRKNFQKKVKKYVELNKHIE-LIHKNLILLEKINTREISIMQV 418
QY 464 FVACLRRLVPPGLWGRHRRERLRTKTFISLGKAKLSQELTW-KMSVRDCAMLR 522
Db 419 ETSAKHFFYFDHE-NIVVLAKLLWIFEDLVLSIRFFVTEQOKSVSKTYIRKNIWD 477
QY 523 SPGVGCVPAEHRLEBILAKFLHWSVYVELRFFYVTEFTOKNRLFFYRKSVWS 582
478 VIMKMSI-ADLKRETLAEVQEKY-EEMKKSGLFAPAGKRLRIPKTTFRPI--MTF--NK 531
QY 583 KLSIGIRQHLKRVQLRSEAEVQHRREARPAALLTSRLRIPKDPGLRPIVNDYVUGA 642
Db 532 KIYNSDRKTKLTNTKLLSHLMKLTKNMFKDPGFVFNVDVDMKKYEEFVCK-WK 590
QY 643 RTRERREKRAELTSRVKALFS--VINYERARR-PGLLGASVLGLDDIHRARWTEVLVRA 699
Db 591 QVGOPKLEFFATMDIEKCYDSVNRKELSTFKTKLLSSDFWIMTAQILKRNKNIVIDSKN 650
QY 700 QDPPPELYFVKVDVTGAYDIPQRLTEVI-AS-IIPQ---NITYCV--RYAVV-Q-KA 750
Db 651 FRKEMKDYFRQKFKIAEGGOYPTFLSVLENQDNLAQKTLTIVEAKQENYFKKNLL 710
QY 751 AHGHVRKA-FKSHVSTLT-DLOPYMRQF-V-AHQETSPLRDVAVIEGSSSLNEASSGLF 806
Db 711 QPVIN-ICQVNYINFNGKFKYQKIGQGLCVSSILSFFYATLEESLGLRDESMPNE 769
QY 807 DVLFRFMCHHA-VRIKRSQVQCGIQPGQISLTSLGSLCYGDMENK-L-F--AGIRRD 860
Db 770 NPVNLMLRITDDYLLITTOENNAVLIEKILNVSRNGFKFNKKLOTSPLSPSFAK 829
QY 861 G----LLRLVDDFLVTPHHTAKTLRLTVRGVPEYGVVNLKRTVVPFVEDEALG- 915
Db 830 YGMSDVEEQNIVQDCWDIGISIDMKTLALPNIN-LRIEGILCTNLNMQTKASMWLK 888
QY 916 -GTAFVQMPAHL-F-PWCGLLDTRILEVQSDYSSYARTSIRASLTFN-RGFRAGRNM 971
889 KLUKSL-MNNITHYFKTTTTFEDFANKTLNKLIFSGYKYMCAKE--YKDHFKKNLAM 945
QY 972 RKLFGVLRKCHSLFLDLQVNSLQTVCTNIVYKILLQAYRFHACVLQLPFHQQVKNPTE 1031
Db 946 SSMIDLEVSIIIVSVTRA 963
QY 1032 FLRVISDTASLCYSILKA 1049

RESULT 6
ID O35432 PRELIMINARY; PRT; 67 AA.
AC O35432;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE TELOMERASE CATALYTIC SUBUNIT (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN [1]
RP SEQUENCE FROM N.A.
RA DRISS R., CLEVELAND J.L.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF029235; G2605903; -.
FT NON_TER 1 1
```

```
FT NON_TER 67
SQ SEQUENCE 67 AA; 8368 MW; E2A06F2B CRC32;

Query Match
Best Local Similarity 5.0%; Score 424; DB 10; Length 67;
Matches 51; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Db 1 FFYVTESTFQKNRLFFYRKSVWSKLSQISGVQRHLERVLRLSEQEEVHHODTWLAMPIC 60
QY 560 FFYVTESTFQKNRLFFYRKSVWSKLSQISGVQRHLERVLRLSEQEEVHHODTWLAMPIC 60
Db 61 RLREIPK 67
QY 620 RLREIPK 626

RESULT 7
ID Q06163 PRELIMINARY; PRT; 884 AA.
AC Q06163;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII COSMID 8543.
GN L8543.12.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA DU Z.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAVELLO A., FULTON L., GAITUNG S., GRECO T., KIRSTEN J.,
RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
RA JOHNSON D., JOHNSON L., LANGSTON Y., LATREILLE P., LE T.,
RA MARDIS E., MENZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA RIFKIN L., TAICH A., TREVASKIS E., VIGNATI D.,
RA WILCOX L., WOULDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA WATERSTON R.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U20618; G662136; -.
SQ SEQUENCE 884 AA; 102663 MW; 1A94320F CRC32;

Query Match
Best Local Similarity 3.7%; Score 314; DB 1; Length 884;
Matches 118; Conservative 122; Mismatches 202; Indels 44; Gaps 37;

Db 271 LSHLSRQSPKER-VLKFTIIVILQKLPOEMFGSKKNGKIKNLNLPLNGYLPDPS 329
QY 448 LVQLLRQHSPPQVQYGVFRACLRRLVPPGLWGRHRRERLRTKTFISLGKAKLSQ 507
Db 330 LLKKLRKDFRWLFIS-DIWFTKHNFNENQLAIC-FISWLFROLIPIOTFFYCTEIS 387
QY 508 LTWKMSYRDCAWLRSPGVCVPAEHRLEBILAKFLHWSVYVELRFFYVTEFT 567
Db 388 STVTIVVF-RHDTWNKLTIPFIVEYKTY-LVE-NNV-CRNHNSYTSNENHSMRIIPK 443
QY 568 FQKNRLFFYRKSVWSKLSQISGVQRHLERVLRLSEAEVQHRREARPAALLTSRLRIPK 626
Db 444 KSNNEFRITAIIPCRGADEEFTIYKENHK-NA-IQPTQKILEYLRNKRPTSF-TKIYSPT 500
QY 627 -PDGLRIVNMDY-VVGARTFRREKRAELTSRVKALFSVLNRYERARRPGLLGASVLGLD 684
501 QIADRI-KEFKQRLKKFNPNVLPYLFMKFDVSKYDSCYDIPRMECMR-ILKDALKNENGFF 558
QY 685 DI-HRAWRTFVLVR-RA-QDPPPELYFVKVDVTGAYDITPQ-DRLTEVIASIIKPNQTYC 740
```

Db 559 VRSQY-FFNTNT-G-VLKLENV-VNASRKPXY-ELYIDNVR-TVHLSNQDVIN---VV- 608  
QY 741 VR-RYAVVQAAHGHVKAFAKSHVSTITDLPYMRQFVAHLQETSPRLDAVVIEQSSLN 799  
Db 609 E---M-EIF-K-T---ALWEDKCYITREDFGSGSSLPVLDVDDLLLEFYSEKAS 658  
QY 800 EASGGLDFELRPMCHAVIRGKSYVQCQIPGSGILSLCSCYCDM-E-NKLF-AG 856  
Db 659 PSQTLKLLADDFLIISTDQOQVINIKKLAMGQFKYNKAN-RDKILAVSSQSDDDTV 717  
QY 857 IRDGLLLRLVDVDFLLVPHLTHAKTELRLTVRGVPEVGVNLRKTVNFPVDEALGG 916  
Db 718 IQFCAM 723  
QY 917 TAFVOM 922

RESULT 8  
ID Q00600 PRELIMINARY; PRT; 234 AA.  
Q00600;  
01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
DE PAROTID 'O' PROTEIN (FRAGMENT).  
GN PRB4.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96108975.  
RA AZEN E.A., AMBERGER E., FISHER S., PRAKOBPHOL A., NIECE R.L.;  
RL AM. J. HUM. GENET. 58:143-153(1996).  
DR EMBL; S80916; G1911492; -.  
FT NON\_TER 1  
SQ SEQUENCE 234 AA; 23656 MW; B5920075 CRC32;

Query Match 1.7%; Score 142; DB 2; Length 234;  
Best Local Similarity 28.2%; Pred. No. 3.44e-05;  
Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;  
Db 92 QGQNQSQG-PPPHGKPERPPP-OGG-NQSHRPPPPP-GKPER-PPPGGNGSQG-P-PP 144  
QY 177 QLGAATQARPPPHGASGRRRLGRCERAWNHSVREAGVPLGLPAPGARRRGGSASRLPLK 236  
Db 145 HPKPEGPPOEGNKSARSPPKPGQPOEGNKPGOPPPPKPGOPPPPGGNGPQOPQ 204  
QY 237 RPRGAPEPERTPVGQSWAHPGTRGSDR-GFCVVSAPPAE-EATSLGALSGRH 294  
Db 205 APPAGKPGPPPPPGGGRPRP 226  
QY 295 SHPSVGRQHAGPPSTSRPRP 316

RESULT 9  
ID Q46612 PRELIMINARY; PRT; 276 AA.  
Q46612;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
DE IS 1222 GENE ORF-A AND ORF-B.  
OS ENTEROBACTER AGGLOMERANS.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC ENTEROBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KLEBERGER, 1983;  
RX MEDLINE; 95255664.  
RA STEIBL H.D., LEWECKE F.M.;  
RL GENE 156:37-42(1995).  
RN [2]  
RP SEQUENCE FROM N.A.

RA STEIBL H.D., SIDDAVATTAM D.;  
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA STEIBL H.D., SIDDAVATTAM D., KLINGMUELLER W.;  
RL PLASMID 34:223-228(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA STEIBL H.D.;  
RL THESIS (1995), UNIVERSITAET BAYREUTH.  
DR EMBL; X78052; G459248; -.  
DR EMBL; X81893; E258949; -.  
SQ SEQUENCE 276 AA; 31718 MW; 0B54A420 CRC32;

Query Match 1.7%; Score 141; DB 9; Length 276;  
Best Local Similarity 33.7%; Pred. No. 4.83e-05;  
Matches 30; Conservative 23; Mismatches 30; Indels 6; Gaps 6;  
Db 42 ITELALERRR-FGYRR-IWQLLRREGLHVNKRVYRLVHLSGLGVKRRRR-RKGLATERL 98  
QY 563 VTETFOKNRLLFFYRKSVWSKLSIGIRQLKRV-QLRELSEAEVQRHREARPAALLTSRL 621  
Db 99 PLL-RPAAPNLTSWSDDFVMDALATGRRIK 126  
QY 622 RFIPKPDGLRPVNMVYVGA-RTFRREK 649

RESULT 10  
ID Q69118 PRELIMINARY; PRT; 296 AA.  
Q69118;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN (FRAGMENT).  
OS HUMAN HERPESVIRUS TYPE 4.  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE;  
OC GAMMAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87284169.  
RA FITZNER A.J., TSAI E.C., STROMINGER J.L., SPECK S.H.;  
RL J. VIROL. 61:2902-2909(1987).  
DR EMBL; M17294; G807646; -.  
KW HYPOTHETICAL PROTEIN.  
FT NON\_TER 1  
SQ SEQUENCE 296 AA; 31393 MW; D17CEF6F CRC32;

Query Match 1.7%; Score 147; DB 11; Length 296;  
Best Local Similarity 38.6%; Pred. No. 6.21e-06;  
Matches 39; Conservative 20; Mismatches 33; Indels 9; Gaps 8;  
Db 91 GPPPTRSGAAQ-RTHRRPPGCRSARNPGCPRTWRR--R-SGAQRHPPPGAGQRRSGP 146  
QY 172 GPPLYQLGAATQARPPPHASG-PRR-R-LGCERAWNHSVREAGVPLGLPAPGA-RRRGGS 227  
Db 147 TGRPAAPAGCTPAAPCGGGAAPVSPGATPHPERGSPAD 187  
QY 228 ASRSLPLKPRPRGAAPER-TPVGGGWAHPGRTGSPD 267

RESULT 11  
ID Q36027 PRELIMINARY; PRT; 574 AA.  
Q36027;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE ACTIN ASSOCIATED PROTEIN.  
GN SPAC4F10.15C.  
OS UNKNOWN.  
OC UNCLASSIFIED.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CONNOR R., CHURCHER C.M.;



RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RA BARRELL B.G., RAJANDREAM M.A., WOOD V.;

RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: 298980; E339884; -

SQ SEQUENCE 574 AA; 59605 MW; 9E84D37C CRC32;

Query Match

Best Local Similarity 1.7%; Score 143; DB 13; Length 574;

Mismatches 42; Conservative 41; Mismatches 53; Indels 10; Gaps 7;

Db 305 AANKRRPPPP-PSRR---NRG-KPIGNGSSNSSLPSPRRSNAAG-SIDLPPQGR 358

QY 180 AATQARPPPHASGPRRLGCE-RAWNHSVREAGVPLGLPAGARRRGGASRSLSPLPKRPR 239

Db 359 SAPPPPPPRAPSAGTGPPLSSSRVSNPP--APPRAIGRSAPALPLGNASRTSTPP 416

QY 240 RGAPEPRTVPVGGSWAHP-GRTRGFSRGCVCVSPARPAEATSLBGAUSGTRNHSHP 298

417 VPTPP-SLPPSAPPSPPLPSAPPSPSLPM 441

QY 299 VGRQHAGPSTSRPPRPWDTPCPVP 324

RESULT 12

ID Q99076 PRELIMINARY; PRT; 316 AA.

AC Q99076;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE HOMEBOX PROTEIN (FRAGMENT).

GN HB9.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE-TONSIL;

RA DEGUCHI Y., KEHL J.H.;

RL NUCLEIC ACIDS RES. 19:3742-3742(1991).

CC -I- SIMILARITY: STRONG WITH OTHER "ANTP-TYPE" HOMEBOX PROTEINS.

DR EMBL: X56537; E30256; ALT\_SEQ.

DR PIR: S16681; S16681.

KW DNA-BINDING; NUCLEAR PROTEIN; HOMEBOX; PHOSPHORYLATION.

FT NON\_TER 1

DNA\_BIND 70 129 HOMEBOX.

DNA\_BIND 112 121 H-T-H MOTIF.

SQ SEQUENCE 316 AA; 34713 MW; 1531E766 CRC32;

Query Match

Best Local Similarity 1.6%; Score 136; DB 2; Length 316;

Mismatches 21; Conservative 20; Mismatches 24; Indels 3; Gaps 3;

Db 202 RPNSTASSDCSVGTGTAIPRGPAI-SPRPSRPAQDRSPARSAPGPAAGPGGAWTH 260

QY 201 RANHSVREAGV-PLGLPAPGARRRGGASRSLSPLPKRPRG-AAPEPRTVPVGGSWAH 258

Db 261 PARPREQA 268

QY 259 PGRTRGPS 266

RESULT 13

ID Q62103 PRELIMINARY; PRT; 317 AA.

AC Q62103;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE PROLINE RICH PROTEIN PRECURSOR.

GN PRP.

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-CD-1; TISSUE-LIVER;

RX MEDLINE; 88273214.

RA ANN D.K., SMITH M.K., CARLSON D.M.;

RL J. BIOL. CHEM. 263:10867-10893(1988).

DR EMBL: M23236; G567232; -

DR MGD; MGI:97773; PRP.

KW SIGNAL.

FT CHAIN 1 15

FT CHAIN 16 317 PROLINE-RICH PROTEIN.

SQ SEQUENCE 317 AA; 31719 MW; 10C84341 CRC32;

Query Match 1.6%; Score 132; DB 10; Length 317;

Best Local Similarity 26.4%; Pred. No. 9.59e-04;

Mismatches 39; Conservative 38; Mismatches 54; Indels 7; Gaps 6;

Db 37 SGSPRPVNGSQGPPPPGQPRP-PQGPVPPGQPPPPGQPPPPGQPPPPGQPPPP 94

QY 180 AATQARPPPHAS--GPRRLGCE-RAWNHSVREAGVPLGLPAGARRRGGASRSLSPLPK 236

Db 95 PPGGQPPRPQGGPPPGG-PQPRPQGGPPPPGQPPPPGQPPPPGQPPPPGQPPPP 153

QY 237 RPRGAAPERTVGGQSWAHP-GRTRGFSRGCVCVSPARPAEATSLBGAUSGTRNHS 295

Db 154 GGQPPRPQGGPPPPGQPPPPGQPPPPGQPPPPGQPPPPGQPPPPGQPPPP 181

QY 296 HPSVGRQHAGPSTSRPPRPWDTPCPVP 323

RESULT 14

ID O02123 PRELIMINARY; PRT; 539 AA.

AC O02123;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)

DE SIMILARITY TO COLLAGENS.

GN W03D2.1.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,

RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,

RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,

RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,

RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,

RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,

RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,

RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,

RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,

RA WILKINSON-SPOAT J., WOHLDMAN P.;

RL NATURE 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA ROHLFING T., WOHLDMANN P.;

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA WATERSTON R.;

RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AF000298; G1947160; -

SQ SEQUENCE 539 AA; 52555 MW; 78FF9A09 CRC32;

Query Match

Best Local Similarity 1.6%; Score 133; DB 3; Length 539;

Mismatches 31.7%; Pred. No. 6.92e-04;

**Matches** 45; **Conservative** 31; **Mismatches** 54; **Indels** 12; **Gaps** 11;

Db 315 AGSPPPPPRGSPPTGSLPPQA-GGSPPPAGT--GSPPPPRQKRQAPERSPP-TGSP 370

QY 179 GAATQARPPPHASGPRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRP 238

Db 371 PTGSP-TGRPPRGPGKSSSESSEREGPRGGPRGGPRKSSSESSEREEPRGPR 429

QY	239	RRGAAPEPERTP-VGQG-SWAHPGTRR-GP-SD-RGFCVWSPARPAEEATSEALSCTR	293
----	-----	--	-----

**D**b** 430 RSPPT-GSPP-TGSPPTGRPPR 449**

QY 294 HSHPSVGRQHHAGPPSTSRPPR 315

RESULT	15
--------	----

RECORDED IN  
ID O41935 PRELIMINARY; PRT; 585 AA.

AC 041935;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)

01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

**HYPOTHETICAL 60.2 KD PROTEIN.**

**GN GAMMAHV .M6.**

OS MURINE HERPESVIRUS 68.

OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; GAMMAHERPESVIRINAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WUMS;

**RX MEDLINE; 97366649.**

RA VIRGIN H.W. IV, LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,

RA DAL CANTO A.J., SPECK S.H.;

RL J. VIROL. 71:5894-5904(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=WUMS;

RA LATREILLE P., WAMSLEY P., WATERSTON R.H.;

RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; U97553; G2317934; -.  
UNIVERSITY OF

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 585 AA; 60160 MW; B35C72E5 CRC32;

Query Match 1.68; Score 135; DB 11; Length 585;

Best Local Similarity 27.6%; Pred. No. 3.59e-04;

Matches	45; Conservative	37; Mismatches	68; Indels	13; Gaps	11;
---------	------------------	----------------	------------	----------	-----

Db 342 PPP-ELGPGSPTSPAPSRAGARIPDLPGPLPSWGPDPRPPPELPGGSPSPAPSRAG 400

QY 173 PPLYQLGAATQARPPPHASGPRRR-L-GCERAWNHSVREA-GVP-LGLPAPGARRRGSSA 228

401 ARIPDLPGPLPSWGPDPRPPPELGP-GSPT-SPAPSRAGARIPPRSRPPPELGP--G 456

QV 229 SRS LPLPKR-PRRGAAPERTPVGOGSWAHPGRTRGPSDRGFCVWSPARPAFEATSI.EG 287

D6 457 SPFFPPPSRPP--PEIGPGSPDI.PGPI.PSWG-PDPPTEFAPSP 496

Qv 288 A1 S C T R H S H P S V C B O H H A C B P S T S P D P D W T O C C P D V A E M Y U 330

Search completed: Thu Jul 30 14:51:34 1998  
Job time : 147 secs.